

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2003, 13.21.54 ; Search time 81 Seconds  
(without alignments)  
1979.018 Million cell updates/sec

Title:	MS-09-769 693-2
Perfect score:	6294
Sequence:	1 MENTOKTVIVTGPLGIYVA ... DELFDLSGPIKHSNTMEM 1203

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 408470 seqs, 133250620 residues  
Total number of hits satisfying chosen parametric

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

1: /SIDS2/qcgcdata/geneseq/geneseqp-emb1/AA1980.DAT.\*  
2: /SIDS2/qcgcdata/geneseq/geneseqp-emb1/AA1981.DAT.\*  
3: /SIDS2/qcgcdata/geneseq/geneseqp-emb1/AA1982.DAT.\*  
4: /SIDS2/qcgcdata/geneseq/geneseqp-emb1/AA1983.DAT.\*  
5: /SIDS2/qcgcdata/geneseq/geneseqp-emb1/AA1984.DAT.\*  
6: /SIDS2/qcgcdata/geneseq/geneseqp-emb1/AA1985.DAT.\*  
7: /SIDS2/qcgcdata/geneseq/geneseqp-emb1/AA1986.DAT.\*  
8: /SIDS2/qcgcdata/geneseq/geneseqp-emb1/AA1987.DAT.\*  
9: /SIDS2/qcgcdata/geneseq/geneseqp-emb1/AA1988.DAT.\*  
10: /SIDS2/qcgcdata/geneseq/geneseqp-emb1/AA1989.DAT.\*  
11: /SIDS2/qcgcdata/geneseq/geneseqp-emb1/AA1990.DAT.\*  
12: /SIDS2/qcgcdata/geneseq/geneseqp-emb1/AA1991.DAT.\*  
13: /SIDS2/qcgcdata/geneseq/geneseqp-emb1/AA1992.DAT.\*  
14: /SIDS2/qcgcdata/geneseq/geneseqp-emb1/AA1993.DAT.\*  
15: /SIDS2/qcgcdata/geneseq/geneseqp-emb1/AA1994.DAT.\*  
16: /SIDS2/qcgcdata/geneseq/geneseqp-emb1/AA1995.DAT.\*  
17: /SIDS2/qcgcdata/geneseq/geneseqp-emb1/AA1996.DAT.\*  
18: /SIDS2/qcgcdata/geneseq/geneseqp-emb1/AA1997.DAT.\*  
19: /SIDS2/qcgcdata/geneseq/geneseqp-emb1/AA1998.DAT.\*  
20: /SIDS2/qcgcdata/geneseq/geneseqp-emb1/AA2000.DAT.\*  
21: /SIDS2/qcgcdata/geneseq/geneseqp-emb1/AA2001.DAT.\*  
22: /SIDS2/qcgcdata/geneseq/geneseqp-emb1/AA2002.DAT.\*  
23: /SIDS2/qcgcdata/geneseq/geneseqp-emb1/AA2002.DAT.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	62.94	100.0	1203	22	AA052633	Varicella-zoster v
2	3016.5	47.9	1196	19	AA072105	HSV-2 strain SB5 C
3	1566.5	24.9	623	19	AA072197	HSV-2 strain SB5 C
4	180.5	21.9	543	19	AA072196	HSV-2 strain SB5 C
5	620	9.9	1152	21	AA053126	Macaca mulatta rha
6	183	2.9	35	20	AA026935	Varicella zoster v
7	181	2.9	35	22	AA073931	Varicella zoster v
8	156.5	2.5	1841	21	AA007562	Protein encoded by
9	123.5	2.0	3411	22	AA073958	Mycelia sterilia c
10	123	2.0	1074	20	AA000206	Enteromomus faeco

11	123	2.0	1.674	20	ABY001878
12	123	2.0	1.674	23	ABY0334077
13	123	2.0	1.674	23	ABY033405
14	122.5	1.9	2.387	21	ABY58254
15	122.5	1.9	2.387	21	ABY53665
16	121	1.9	1.069	21	ABY811664
17	120.5	1.9	2.597	21	ABY536968
18	120.5	1.9	2.597	21	ABY879938
19	118.5	1.9	1.076	22	ABY027075
20	118	1.9	1.784	22	ABY034772
21	115	1.8	1.150	14	ABY839505
22	114.5	1.8	1.456	20	ABY010703
23	114.5	1.8	1.456	21	ABY070466
24	114.5	1.8	1.456	21	ABY115878
25	114	1.8	1.458	22	ABY060616
26	111	1.8	4.630	7	AAW196228
27	111	1.8	4.630	21	AAW771777
28	110.5	1.8	1.532	22	ABE685858
29	110.5	1.8	3.070	22	ABG8119970
30	109	1.7	1.1096	22	AAE101222
31	108.5	1.7	1.451	22	AAH358766
32	108	1.7	2.556	22	AAH305676
33	106.5	1.7	7.74	16	AAH863737
34	106.5	1.7	7.94	22	ABG071777
35	106.5	1.7	3.298	23	ABG054330
36	106.5	1.7	3.567	14	ABR444331
37	105	1.7	7.74	18	AAW11700
38	104.5	1.7	1.471	19	AAW45585
39	104.5	1.7	1.515	23	AAU10533
40	104.5	1.7	3.530	22	AAW55586
41	104	1.7	7.74	13	AAK221222
42	104	1.7	7.74	16	AAH939848
43	104	1.7	8.21	13	AAK21080
44	103.5	1.6	1.750	22	ABR21522
45	103	1.6	7.74	14	ABR4081

## ALIGNMENTS

RESULT	1
AAM52633	
ID	AAM52633 standard; Protein; 1203 AA

AC AAM52633;

DT 18-FEB-2002 (first entry)

DE Varicella-zoster virus (VZV) ORF29p protein

KW ORF29p; open reading frame 29p protein; Varicella-zoster virus;

drug delivery, gene delivery, fusion protein, protein secretion

OS Human herpesvirus 3

PN W0200154709-A1.

PD 02-AUG-2001

PF 25-JAN-2001; 2001WO-US02500.

PR. 25-JAN-2000; 2000US-0177901

PA (NYCO ) UNIV COLUMBIA NEW YORK

PI Silverstein S, Annunziato P, Gershon A, Lunqu U

DR WPI; 2001-488744/53

XX  
DT + 70004 E3+

PT Composition of matter for delivering desired compounds into eukaryotic cells, comprises Varicella-Zoster virus open reading frame 29p protein either bound to an agent or operably attixed to lipid-soluble group.

Enfermoccus laevis  
E laevis EF094 F  
F facialis EF102 F  
Drosophila melanogaster  
Mechanical stress  
Human OPRC EF170C  
Mechanical stress  
Pat verp Patius I  
Novel human diatom  
Novel human diatom  
C protein fragment  
Melanoma associated  
Human p53 taget in  
Human peroxidase  
Cephalosporin C ac  
Streptomyces venosus  
S. venezuelae vsp  
Drosophila melanogaster  
Mycobacterium tuberculosis  
Streptomyces nebulosus  
Human protein seqs  
A splice variant  
Cephalosporin C  
Novel human diatom  
Human dactylos protein  
erya region polypep  
TTP168try cephalosporin C  
Calcium independent  
Kat C1R1-1 variant  
Human protein seqs  
Sequence encoded  
Cephalosporin C  
flg receptor protein  
Novel human diatom  
precursor mutant

XX Claim 1: Fig 6: 68pp: English.

XX This sequence represents the Varicella-Zoster virus (VZV) open  
 XX reading frame 29p (ORF29p) protein. Although ORF29p is the major DNA  
 XX binding protein of VZV, the invention is based on the discovery that the  
 XX VZV ORF29p protein can readily enter and exit eukaryotic cells.  
 XX Accordingly, the invention relates to a composition for delivery of an  
 XX agent into a eukaryotic cell, comprising the VZV ORF29p protein bound to  
 XX the agent to be delivered. The agent delivered may be a polypeptide, a  
 XX polynucleotide, or an organic compound. The invention also relates to  
 XX a composition in which the ORF29p protein is linked to a lipid soluble  
 XX group that permits the protein to be anchored to a lipid membrane, and  
 XX lipid vesicles containing this composition. The invention additionally  
 XX relates to the use of the ORF29p protein as a facilitator of secretion of  
 XX a desired protein, in which the desired protein and the ORF29p protein  
 XX are recombinantly expressed as a fusion protein. The invention further  
 XX encompasses nucleic acids encoding the VZV ORF29p protein and their use  
 XX in detection of VZV ORF29p nucleic acids, and a monoclonal antibody  
 XX against the ORF29p protein. Compositions of the invention are useful for  
 XX the delivery of a prophylactic or therapeutic agent to a eukaryotic cell,  
 XX particularly a human cell. Examples of agents that can be delivered to a  
 XX cell include proteins such as insulin, factor VIII, factor IX, and  
 XX proteases; polynucleotides (e.g., for use in gene therapy); and organic  
 XX compounds such as vitamins and a wide variety of pharmaceuticals (e.g.,  
 XX anti-infective, and labeled or immunosuppressive drugs).

XX Sequence 1203 AA:

Query Match: 100.0%; Score 6,294; DB 22; Length 1203;  
 Post Local Similarity 100.0%; Pred. No. 0;  
 Matches 1203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

UY 1 MENTOKTVTPDGLGVYACREVDJDEEISPLAARSTDSIALPLMNLIVETFTS 60  
 DB 1 MENTOKTVTPDGLGVYACREVDJDEEISPLAARSTDSIALPLMNLIVETFTS 60  
 UY 61 SLAVVSGAATGTLGAGGTLKLTSHFYPSVYFPGGKIVLPSSAAPNLTRACNAAREP 120  
 DB 61 SLAVVSGAATGTLGAGGTLKLTSHFYPSVYFPGGKIVLPSSAAPNLTRACNAAREP 120  
 UY 121 GPSKVGPPVKAIVETTGAFICITRLCEPENTILLYLVATLKEAFMGNVFLHYGLDI 180  
 DB 121 GPSKVGPPVKAIVETTGAFICITRLCEPENTILLYLVATLKEAFMGNVFLHYGLDI 180  
 UY 181 VHHNGDVRILFLFYVULFMPVNVKLVPDPNTHRSICGCVYVPPVNTGLCHLHNC 240  
 DB 181 VHHNGDVRILFLFYVULFMPVNVKLVPDPNTHRSICGCVYVPPVNTGLCHLHNC 240  
 UY 241 VIAPMAVALKRVNNTVAVAGAAHLAFDENIEGAVLPDPITYTEOSSSGTTARCARRN 300  
 DB 241 VIAPMAVALKRVNNTVAVAGAAHLAFDENIEGAVLPDPITYTEOSSSGTTARCARRN 300  
 UY 401 DVNSTRKPSPTSEFEERASIMAMATLAAEVFNTGIVEETPTDKKEMPMETGMEETLP 360  
 DB 401 DVNSTRKPSPTSEFEERASIMAMATLAAEVFNTGIVEETPTDKKEMPMETGMEETLP 360  
 UY 461 KLMALASTYARAVAGVIGAVPSPNSALYTFEVDNSGMTAKKGGPSPSNRYQAPGPHL 420  
 DB 461 KLMALASTYARAVAGVIGAVPSPNSALYTFEVDNSGMTAKKGGPSPSNRYQAPGPHL 420  
 UY 461 KLMALASTYARAVAGVIGAVPSPNSALYTFEVDNSGMTAKKGGPSPSNRYQAPGPHL 420  
 DB 461 KLMALASTYARAVAGVIGAVPSPNSALYTFEVDNSGMTAKKGGPSPSNRYQAPGPHL 420  
 UY 421 AANPOTDRKHIVLSSSGTSSNTEFEVDLALICGGAFLALFLYLERGDMGATGGH 480  
 DB 421 AANPOTDRKHIVLSSSGTSSNTEFEVDLALICGGAFLALFLYLERGDMGATGGH 480  
 UY 481 GHALKYVTGTFESETPGSLCEKHTPRVAAHTTVHRLRQMPREGQATROPIGVEGTMNSQ 540  
 DB 481 GHALKYVTGTFESETPGSLCEKHTPRVAAHTTVHRLRQMPREGQATROPIGVEGTMNSQ 540  
 UY 541 YSLDTPGLGVAAYLLIKRKQDQTEAAKATIMQDYRATLERFLIDLEOERLLDGGACSSSE 600  
 DB 541 YSLDTPGLGVAAYLLIKRKQDQTEAAKATIMQDYRATLERFLIDLEOERLLDGGACSSSE 600

UY 601 GLSSVIVHPTFPPLIDLEPARIEGTTOPMKVLEVETGVKIRKGLSEATHSMALTEDPY 660  
 DB 601 GLSSVIVHPTFPPLIDLEPARIEGTTOPMKVLEVETGVKIRKGLSEATHSMALTEDPY 660  
 UY 661 SGAPCPITNPLVAKTILAVVQDLALSGCHCVYGVGVGCAKPKKQVQVYVKKRVQVLFNG 720  
 DB 661 SGAPCPITNPLVAKTILAVVQDLALSGCHCVYGVGVGCAKPKKQVQVYVKKRVQVLFNG 720  
 UY 721 GFISTRTITTSKGVASAPNPLISQVAPAGKTFHGLAVSVVVKDVRKKNVYVSGN 780  
 DB 721 GFISTRTITTSKGVASAPNPLISQVAPAGKTFHGLAVSVVVKDVRKKNVYVSGN 780  
 UY 781 CTNLSAARARLVGLASAVYRUEKRVDMGLGALFLLKQFHLFPKMPNKSPPQW 840  
 DB 781 CTNLSAARARLVGLASAVYRUEKRVDMGLGALFLLKQFHLFPKMPNKSPPQW 840  
 UY 841 FWTLLQKQNPAPKLTHEETTTAAVKKPTEBYAALNFINPPTGICGLAOFYMANILK 900  
 DB 841 FWTLLQKQNPAPKLTHEETTTAAVKKPTEBYAALNFINPPTGICGLAOFYMANILK 900  
 UY 901 YCHSGLVINTLTSITGARRKPPSSVLHMKKQVTSADIFEOAKALLKTEKNEPELW 960  
 DB 901 YCHSGLVINTLTSITGARRKPPSSVLHMKKQVTSADIFEOAKALLKTEKNEPELW 960  
 UY 961 TIAFTSTHIVKAAKMPVNVLSISIKYHSAANNEVFAGNMSSLNGKKNVPLETTFD 1020  
 DB 961 TIAFTSTHIVKAAKMPVNVLSISIKYHSAANNEVFAGNMSSLNGKKNVPLETTFD 1020  
 UY 1021 KTRRFIICAPRGGFCGVTPGSSNRPETTSIQVKGILIVSGAMVQIATATVYVAVGAR 1080  
 DB 1021 KTRRFIICAPRGGFCGVTPGSSNRPETTSIQVKGILIVSGAMVQIATATVYVAVGAR 1080  
 UY 1081 AQHMAFDWMLSLTDDLEFLANDLELHDOILQTEFTWTVGALAAVAKLDEKKTAGIGCT 1140  
 DB 1081 AQHMAFDWMLSLTDDLEFLANDLELHDOILQTEFTWTVGALAAVAKLDEKKTAGIGCT 1140  
 UY 1141 PTMLAFNFDSCGESHDTSVNLISGNSIGSTVPGIKRPPEDDELFDLSGTPKKNIT 1200  
 DB 1141 PTMLAFNFDSCGESHDTSVNLISGNSIGSTVPGIKRPPEDDELFDLSGTPKKNIT 1200  
 UY 1201 MEM 1203  
 DB 1201 MEM 1203

RESULT 2  
 ID AAW72105 standard; Protein; 1196 AA.  
 AC AAW72105;  
 DF 18-DIC-1998 (first entry)  
 DE HSV-2 strain SD5 Contig ID 12 ORF#2 protein.  
 KM HSV-2 strain SD5; immunological response; Antigen; 1203; 1203; 1203;  
 KM antiviral identification; viral protein inhibitor.  
 OS Herpes simplex virus type 2.  
 PN W09820016-A1.  
 PD 14-MAY-1998.  
 PP 31-OCT-1997; 97WO-US240016.  
 PR 09-JUN-1997; 97US-0049018  
 PP 04-NOV-1996; 96US-0030279.  
 PA (SMK) SMITHKLINE BEECHAM CORP.  
 XX Chan JY, Dabrowski-Amarel CE, Delvecchio AM, Dillon SB;  
 PI Esser KM, Leary JT;

XX WPI: 1998-286847/25.  
DR N-PSDB: AAV62156.

XX Herpes simplex virus type 2 sequences - useful in, e.g. prevention  
PT and treatment of infection or inducing immunological response in  
PT mammal

XX Claim 10; page 83-84; 748bp; English.

XX This sequence represents a Herpes simplex virus type-2 (HSV-2) protein  
CC sequence of the invention. This sequence was isolated from a HSV-2 strain  
CC SB5 (deposited as ATCC Vb-2546) DNA fragment designated Contig ID 12.  
CC The proteins can be used for the treatment or prevention of disease, to  
CC induce an immunological response in a mammal or to identify inhibitors,  
CC activators or novel antivirals. Antagonists of the proteins can be used  
CC to inhibit a viral polypeptide. The DNA sequence or a vector containing  
CC it can also be used to induce an immunological response in a mammal

XX Sequence 1196 AA;

Query Match 47.9% Score 3016.5; DR 19; Length 1196;

Best local similarity 49.1% Freq. No. 1.7e 284; Indels 41; Gaps 16;

Matches 596; Conservative 211; Mismatches 367;

4 TUKTVETGPGVYV--ACRVEDDLEBEISFLAARSTDPLALPLMNLTVKFTFS 61  
1 TTTTVKVPKPGVYVYGNACPAEGLEL--LSLSAASGDAVAAPLIVGLVESGFHAN 64  
62 LAVVAGARTTGLAGGITKLTSTHTFVFPVFGKIVLPSSAANLTPACNAAREPG 121  
65 VAAVVSRTTGTGTAVALSKLMPSTSPSVYVHGGRHLPATQANLPLTERAPRHG 124  
122 FSRGCGTVDGAVETTCATCTRLCTLEPNTLLYLAVTALFKAAYMCVWPLHYGGL 181  
125 FSDYAPRCODLKHETTGALCERLGLDPRALLYLITGFRBAVCISTFHLGMDKV 184  
182 HINRGDVTIRPLEPVGLFMPDVNRLVPRPNTHTSTGEGFYVPPRYNTGICHLIHPV 241  
185 TIGDAEVNHRIPVPLQMPDPDSKVADPPNCHRSIGENFVPLDFNRLARLLFEAV 244  
242 IAPMVALKRVNVAVKGAHLAEDENEHVAVLPDITVTFOUSSSITTAKAKKND 301  
245 VGRAVALLARVNDVARRAHLAEDENEGAALPADITFTFEASQG--KPRGAR--- 299  
302 VNSTKSPSGCFERRIASIMADTALHAENVIFNTGIVEFTPDIKEMWFIQMEGTLDP 361  
300 --DAGNKGPAAGFEGORLASVMAGDALALESTVSMVDPERPDDITWPLLEBOETPAR 357  
362 LNALGSYARVAGVIGAMVFSNLSALYLTVEEDSGMTEAKDGSGPSENFOTOGPHLA 421  
368 AGAVGAYLAKRAGLGVAMVSTNSALHLEVDAGPADPKDSK--PSFYRFLVPTHYA 416  
422 ANPDTPDQHNV-----SSQSTGSSNTSEFSDYALLCGSGARPLAPLPLEGCDAG 474  
417 ANPDREHNVVPGEGRTAPLVGTO--EFAGENHLMACGSPALLAMLFLEKCDG 475  
475 AVTGGHG--DALKYVTGTPDSEIPCSI/CEKHTRPVCAHTTVHRLQMPRGQATROPICV 533  
476 VIVGREGMVFYKVVADSGIDVPCNLCTFETRIACVHTLMLRLARHPRKFAAPALGIV 545  
534 EGTMSUYSTCTGFGNAYAYLLIPKIKVTEAKATMCTYATLDELFTLEGFLLIP 543  
536 FCIIMSAISDCIVLQNAHSAIKR--AKDSENTKTIIMDEYTKATEKVMALAEALQYVQ 544  
594 GAPSSEELSSVIVDPFRITLPIRARIEQTTQPMFLVETPRYKIPREGSEATHS 653  
595 AVPTALGRIETIIGTEALHTVANNIKOLVREVEQMLNLLIEGRNFKTRDGLAEANHAM 654  
654 ALTHPPYSACFCTINILVKKHILAVQVIALSGCHGVYVGGGVKKNRNPQVPLKKE 713  
655 SLSDPYTGTPCLDLKLKSNLAVYQDLALSCVTHSVFASQSVDEIRNPNQFQVPLPRK 714

714 FVDLENGFISTRITVLTSEG--PVSAPNPLISQULAPAR:PTFLGDLAPVSEVETPIPK 772  
715 VADLENNGLSAKTILVALISEGALICAPSLITACQTPAPASSPFGDVARVITGLPKELRK 774  
773 NRVVSGNCTNISEAKARIVGLASAYQORKEKVMHGAICHLIKOFHQLLPPKMPIN 832  
775 SRLVFGASANSSEAKARIVASLOSAYQKPKDVRDITLPLGLLKOFAVIRPNNKPKPG 834  
833 SKSPNQWHTLLQKQWALUKITHEIITIAAVKFTETRYAALNINLPPTICGLAQ 892  
835 SNQPNQWHTALQKQWALUKITHEIITIAAVKFTETRYAALNINLPANNVSLIANY 894  
893 YMANLLKUCDSUYLINTLSITGAKKRPDPSVLMIRKIVTSAADLETAKALLEK 952  
895 YMANQILKRCDSHYEINTLITAVIAGSRPPQVQAAANAPQ---GAGIGLACARALMDS 951  
953 TENLEPLMTATFSLHLYAANQRPVYVIGISIKYHAGACNNRVQACNMSGLOCKN 1012  
952 LDAHGAWTSMFASCKLILPVMARPMVYLGISIKYHAGACNNRVQACNMSGLOCKN 1011  
1013 VCPLEFDRTRRPLIACPRGFCIPVTGPGSSGRRFTTISDOVGIIVSGAWQALAYAT 1072  
1012 ACPLILFDRTKRFVLACPRAGFYCAASSJGGAHNSLICHQLKGLIABGAAVASSVYA 1071  
1073 VYRAVGAQAQHAFFDMLSLTDEFLARDELELHDQIQTLETPTVEGAL---EAVKI 1128  
1072 TVKSLCPRTQQLQIDFWLALILEDEYSELSEMMFEPTTALREGHGEMSDALAEVAHFAAL 1131  
1129 LDEKTIAGQDEITNLAPNUSCEPSPHUTISVNLNLSGNSISSTIVGLKKRPEDDRPLD 1188  
1132 VSQLGAAGE-----VFNFSGDFGDEDDHIAASFSIAAA--AGAAGVARRAHRGDDPE 1182  
1189 LSGIPKIKHGNITMEM 1203  
1183 RGPPEKK-NPRTIDM 1195

RESULT 3  
AAW72197  
ID AAW72197 standard; Protein; 623 AA.  
XX  
AC AAW72197;  
DT 13-JAN-1999 (first entry)  
DE HSV-2 strain SB5 Contig ID 15 ORF#12 protein.  
XX  
KW HSV-2 strain SB5: immunological response induction; therapy;  
KW antiviral identification; viral protein inhibitor.  
XX  
OS Herpes simplex virus type 2.  
XX  
PN MO9820016-A1.  
XX  
PD 14-MAY-1998.  
XX  
PF 31-OCT-1997; 97MO-US20016.  
XX  
PR 09-JUN-1997; 97US-0049018.  
PR 04-MAY-1996; 96US-0030279.  
XX  
KA (SMIK ) SMITHKLINE BEECHAM CORP.  
XX  
VI Chan JY, Dabrowski-Amara CE, Delvecchio AM, Dillon SH;  
PI Esser KM, Leary JJ;  
XX  
DR WPI: 1998-286847/25.  
XX N-PSDB: AAV62156.  
XX  
PT Herpes simplex virus type-2 sequences useful in, e.g. prevention  
PT and treatment of infection or inducing immunological response in  
PT mammal

ES Claim 10; Page 123; 748bp; English.

XX This sequence represents a Herpes simplex virus type 2 (HSV 2) protein  
 CC sequence of the invention. This sequence was isolated from a HSV-2 strain  
 CC SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 15.

CC SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 15.

CC Based on homology, this sequence is a major DNA-binding protein.  
 CC The proteins can be used for the treatment or prevention of disease, to  
 CC induce an immunological response in a mammal or to identify inhibitors,  
 CC activators or novel antivirals. Antagonists of the proteins can be used  
 CC to inhibit a viral polypeptide. The DNA sequence or a vector containing  
 CC it can also be used to induce an immunological response in a mammal.

XX Sequence 624 AA:

Query Match 24.9%; Score 1566.5; DB 19; length 623;

Best Local Similarity 50.0%; Pred. No. 2,90-143;

Matches 113; Conservative 100; Mismatches 192; Indels 21; Gaps 9;

4 TUKTVTPTGLAYVY ACRYEDLLEELFLAANSTSDLLALMLNLTVEKFTSSL 62

7 TTTTAVKPHGCMGYVYKATTAEGLEL-LSLSARSGADVAVALVGLTVESSCEAVV 64

65 AAVSGARTTGACACILITKLTSHYVYFPHCKHVLSSAALNIPACMAARHRCF 122

65 AAVSGARTTGACACILITKLTSHYVYFPHCKHVLSSAALNIPACMAARHRCF 122

123 SPQGPVPCAVETPAZLCTHATFPENTLYVVTALPKFAVPMCVNPHVGGDPIVH 182

125 SHVAPRKHCKHETGDMALCELDIDPDAALYVITPCCKRFAVGLSNTECHLQMDKVT 184

185 ENHIVITETITVOLEHMDVNRKLVDFEFTNHSTGRGTVPFTFYNTGCTHILHCVI 242

185 TGDAAVHRHVPVYVGLMPHDSFVLAIPPNNHRSICFNPNVLCFFPNNTAKLFEAAV 244

243 APMVAVALRVNVTAAAKCAHLAFPHNCGAVLPDDIYTFYVSSSSSTTARCARNDV 302

245 GFAVALAPVAVVAVRAAHILAFLENHFCALLPADITFTFAEAVSG- KPGFAP- 298

101 NSTSKPSTSGCKERRKSLMAADTALAAVLPNTGCTEETPDIKKVMFICMECHLRL 362

209 IACNKGACGCEHMLASVMAADAAALLESIVSAVFPDIPDDITWMLDQFTTAARA 357

363 NALSYTARVAVIGAVGESPASLYTEVEDSGTEAKDGGPESFNHYGFAPSHLAA 422

368 GAVATVYLAANAIVGAMVSTNSALHILEVDIAGADIKDHSK-PSYKRPFLVPTHYAA 416

423 NVQDIDRCHVL SSOSTGSSNHPESVDYALICGQAPILLAKLFLYLERCDAGA 475

417 NPQLDREGHVVVREKRPATLAVGCTO-EFAGEHILAMLYGFSPLDLAKMLEYLERCDGV 475

476 FEGGNG DAIKYVGLIHSEIENGSCGKHTRVCAHTTVHRLKORMPHRCQATKQPIGV 534

476 TVGCKHMLVFRVVAISVQTDIVGNCNCTFETRHACAHITLMIRAKRHPKASAKARACIGVF 535

545 GTMNGVSNVSTFNVAAYVILIKRPGQTEAKATMCTEATLEPFLDEGDELLDPS 594

546 GIMNSAVSICVAVTNYAAPSALKR-ADGSENRITIMQETVAAALIKRYVAHLALQYVIGA 594

595 APSSSGTSSVAVVAGTFRKRLDITIK 620

595 VPTAARLETTITGKALHTVNNIK 620

RESULT 4

AAW72196

AAW72196 standard; Protein: 543 AA.

AAW72196

13 JAN 1999 (first entry)

HSV 2 strain SB5 Contig ID 15 ORF#31 protein.

KW HSV 2 strain SB5. Immunological response induction; therapy;  
 KM antiviral identification; viral protein inhibitor.

OS Herpes simplex virus type 2.

PN W09820016-A1.

PD 14-MAY-1998.

PF 31-OCT-1997; 97WO-US20016.

PR 09-JUN-1997; 9/US-0049018.

PK 04-NOV-1996; 96US-0040279.

PA (SMK ) SMITHKLINE BEECHAM CORP.

P1 Chan JY, Dabrowski Amarel CE, Delvecchio AM, Dillon SH;

DR WPI: 1998-286847/25.

DR N-PDBH; AAV62176.

PT Herpes simplex virus type-2 sequences useful in, e.g. prevention

PT and treatment of infection or inducing immunological response in

PS Claim 10, Page 122-123; 748bp, English.

CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein

CC sequence of the invention. This sequence was isolated from a HSV-2 strain

CC SB5 (deposited as ATCC VR-2546) DNA fragment designated contig ID 15.

CC Based on homology, this sequence is a major DNA-binding protein.

CC The proteins can be used for the treatment or prevention of disease, to

CC induce an immunological response in a mammal or to identify inhibitors,

CC activators or novel antivirals. Antagonists of the proteins can be used

CC to inhibit a viral polypeptide. The DNA sequence or a vector containing

CC it can also be used to induce an immunological response in a mammal.

XX Sequence 543 AA:

Query Match 21.9%; Score 1380.5; DB 19; length 543;

Best Local Similarity 48.7%; Pred. No. 3,48-125;

Matches 271; Conservative 101; Mismatches 165; Indels 19; Gaps 7;

653 MALTRVPPSGACPIITNPLVKKRTHLAVDVIALSUCHVYVCGUQVGRNPRNPPVLR 712

1 MSLSIDPPYCGCPDQLLARRSNLAVDIALSOCHGVAGQSVGRNPRNPPVLR 60

713 RVDFENGGFISTRTSTVTSFG-PVSAPNPTLGPAPAPPTFGDIAVSEVEIDIVP 771

61 RMDLPPNNGFLSAKTITVALSDEGALCALSTACOTAPRKSSTFGVAVVTLGPKRLKV 120

772 KRVVSCNGCTNLSAARALVGLASAVOROKKVMIMGALCFLLKOPHGLPFGMP 831

121 KSRVLFAGSANAASAAKRAVASTQSAVCKEDKVIDLIGPFLKCPHAAVEFGNPKP 180

832 NKSVPNQPWFULLQORNPADKLTHERITTAAYKRTFETKALNFTNLPITGHLAQ 891

181 GSNQNPQWFWALDQNPAPLSPHDLITAFIKPSLIDGALNFINMANNVSEIAM 240

892 EPMANLILKYDHSOYLINTLSITFGARRPDPSVLIMIKDVTSAADIELOKALLE 951

241 YMANQILKCYCHSYTFINTLFAVLAQSKRPYCYGAAAMAVQ---GGAGLPAKARALMD 297

952 KTEHNPRLMTTATSTHLVRAAMNPPVAVGISTSKYHGAAGNNVPAQAWSDI NGSK 1011

298 SLDAHGAWTSMEASCNILRPVMAARPMVVLGLSTISKYVGAGNDVRVQAGWASLIGSK 357

1012 NVCPILFTDRIRKPLIACRGGFICVIVGPPSGNRRHTLSDVGRGIVSGGAMVQALIVA 1071

358 NCPPLIFDPRKPVYACRAGSFVCAASSLUGGAAHHSICVLPKGLIARCGAAVASSVAV 417

1072 TVVRAVGAHQHMAFDWMLSTDEFLARDELHIDQIIQIETFWTVEGAL---EAVK 1127



DB 418 AAVKSLGPKHIOUQJEDWALIEDEYLSHEMMFTTRALRGHGEWSTDAALVAAHAA 477  
 QY 1128 IDEKRTAGSGETPINLAENFSCFESHDTSTNVLANISSNLSKSVPLKKEPDEDEL 1187  
 DB 478 IYSQJGAACH-----VNFQDQKQJHJHAAHSPGLAA--AGAAQVAKKAFHJDDPP 528  
 QY 1188 DLSGIPKHKNTTMM 1203  
 DB 529 G-EGPPEKK-DLTLDM 542  
 RKSUIT 5  
 AAB53126  
 ID AAB53126 standard: Protein; 1132 AA.  
 AC AAB53126:  
 XX 28-FEB-2001 (first entry)  
 DB Macaca mulatta rhadinovirus 17577 RRV ORP6 protein SEQ ID NO:4  
 XX  
 XX Macaca mulatta rhadinovirus 17577: RRV; rhesus macaque rhadinovirus;  
 KM genome; Kaposi's sarcoma-associated herpesvirus; KSHV; interleukin 6;  
 KM IL-6, macrophage inflammatory protein; MIP; diagnosis, vaccine,  
 KM cytostatic; anti-HIV; gene therapy; infection; Kaposi's sarcoma,  
 KM lymphoproliferative disorder; B-cell hyperplasia; lymphadenopathy;  
 KM splenomegaly; hypergammaglobulinemia; autoimmune haemolytic anaemia.  
 XX  
 XX Macaca mulatta rhadinovirus 17577.  
 OS  
 XX WO200028040-A2.  
 PN  
 PD 18-MAY-2000.  
 XX  
 XX 05-NOV-1999; 99WO-US26260.  
 PP  
 XX 06-NOV-1998; 98US-0107507.  
 PR  
 XX 20-NOV-1998; 98US-0109409.  
 XX  
 PA (UNIV OREGON HEALTH SCI  
 PI Wong SM, Axthelm MK, Seattles RP;  
 XX  
 DR WPI: 2000-376552/32.  
 PT  
 PT New rhesus rhadino virus for producing non-human primate model useful  
 for testing potential treatments and efficacy of the candidate vaccine  
 for conditions associated with RRV infection -  
 XX  
 PS Claim 5; Page 123; 141pp; English.  
 XX  
 XX The present invention describes a novel rhesus macaque rhadinovirus  
 CC called macaca mulatta rhadinovirus 17577 (RRV). AAC64754 represents the  
 CC RRV genome sequence, and AAB53123 to AAB53204 represent the proteins  
 CC encoded by the genome sequence. The present invention also specifically  
 CC claims the individual open reading frame (ORF) nucleotide sequences from  
 CC the genome which encode the individual proteins, but these sequences are  
 CC not given. A non-human animal infected with RRV can be used for testing  
 CC the efficacy of drug in the treatment of condition associated with  
 CC infection with RRV such as Kaposi's sarcoma, lymphoproliferative  
 CC disorders, B-cell hyperplasia, lymphadenopathy, splenomegaly,  
 CC hypergammaglobulinemia or autoimmune haemolytic anaemia, by  
 CC administering the drug to a immuno-compromised non-human primate  
 CC preferably rhesus macaque monkey obtained by as a result of infection  
 CC by Simian Immunodeficiency Virus (SIV). RRV is useful for producing  
 CC non-human primate model for testing potential treatments for conditions  
 CC associated with RRV infection. It is also useful for testing the  
 CC efficacy of the candidate vaccine against RRV infection or conditions  
 CC associated with its infection by administering the vaccine to the  
 CC subject capable of infection with RRV, inoculating the subject with RRV  
 CC and observing the effect of vaccine. AAC64755 to AAC64765 and AAB53205  
 CC to AAB53213 represent sequence used in the exemplification of the

CC present invention.  
 XX  
 SQ Sequence 1132 AA;  
 Query Match 9.98; Score 620; DR 21; Length 1132.  
 Host Local Similarity 22.68; Fred. No. 2.7e-50;  
 Matches 290, Conservative 212, Mismatches 506; Indels 274, Gaps 50;  
 QY 1 MENTQKTVTVPTGPGVYVACPVEDLDLEISFLAASSTSDALPLPMLNLTVEKFTTS 60  
 DB 11 LEDNGS-RAPIGACGVYVAYSQODEPFAASTLGNPSSGVEFSLPILGLVHEHFFPL 69  
 QY 61 SL-AVVSAGATTGLACAGITLKLTTHSHFVSVVPHQKAVLPSSAANPITRACNAAR 119  
 DB 70 TVKAAVKKVDTTLA-----VKVTC--FIREVIVFINASLPFVFDGTGELNECEAFAL 122  
 QY 120 FGFSSRC--VQSP--VQAVETTGAEICTRGLEPENTILYVTAJLKE-----AV 166  
 DB 123 FGTOPIENGCPHPSIWNPLE-----CPQLP-DKDEMPICGVVTEGKREKLMKQCLVPAV 175  
 QY 167 FMCNVLHWGSLDVIHINSHDVIPIPLFVQLFMPVNPVLPVPPFNTNHSISGEFVYPT 225  
 DB 176 FLYVQ-----VLASQVAFKVPFLYDEDLFAPIHGRM-----P 207  
 QY 227 PRYNTGLCHLHCCVLAEMAVALKVKNVTAVAGAAHLADLNHGAVALPPDITTYTQS 286  
 DB 208 RFYHKQVSAVLVDSLFSTSLQALRLKLVTAIVHAIEKQFMJHYKIAKI---VQAKJF-- 262  
 QY 287 SSSGTTAPGAPPNVNSTKSPSGGFEPRLASTMAADTALAEVIFNTGIYEERTPI 346  
 DB 263 -----STLPKTTDSSHMIVDSVAELALSYGCH-----LCEHODA 300  
 QY 347 KE-----WEMPTGMEHTLPLNALASVTARVAVSVIYAAWFSFNSALYILEVELSMTFA 400  
 DB 301 CELLNVDWPIFDGCDSPPEARVNALEBWSAQVHVAGQLFAANSVLYLTKVQKQAPRG 360  
 QY 401 KDSGPGSPFNKFTYQKAPPLAANVQDTRDCH-----VLSQSSTGSSNTSPDYIALICG 455  
 DB 361 K--GDVWVNSFPLQHLQGLNATIKENSGEAFKQVPSNALDSSFTPY--HLAAVAS 415  
 QY 456 FGAPLLARLFLFLEPRIDAFAGFTGSHSDALKYVTSIFPSELPSICEHTRPVCNHTVHR 515  
 DB 416 FSPHLLAKLYYWGFIQHHSSSTNGAFNMHYGTATANSKMTCHCHNTPATCINTLPYR 475  
 QY 516 LRQMPREFGVATRG-LIVGFTGNSQYSDCDPLGANVAYLILKRPDGT-EAAKATMDT 573  
 DB 476 LKDRFPAVTTQRRDDPVYVYTAGT--FNDEILIGNFASFRDREDGDPADENHPKYTWQL 534  
 QY 574 YRATLEPLFLDQLQERLLDGCAPCSSSGLS-----SVLVDPHTPRILLDTLARIHOT 626  
 DB 535 CQTVTEKL-----SALGITEDHDHNVNLTNTQSLFVRKGDISTVIGE 578  
 QY 627 TTQFMKVLVETRDYKIREGLSEATHSMALTFDPSGAFPTTNFLVPRTHLAAVVDLALS 686  
 DB 579 VMFVNSMIK-NMNPFRHKSVAHHILOFCPCNVWQAPCAVFLNLYKSLIMI IQDILCP 637  
 QY 687 QCRVVRVYQ-----QVEGRNFRN-----QFQVLPFRFVDLF 718  
 DB 638 --YCMITEDNPMWGLPSEMLKMHFYLMTNFKAA-LDRGVLTGTLKTVHPMPDFDF 695  
 QY 719 -----NCGFISTSHSIWTLSEGPSAVNPPIHGOAPAGKIFLQGLAVASVGLIDIVKN 773  
 DB 696 DTAAGSKGLMAFRKMYR-----IAAMNVFRSSTIKN 724  
 QY 774 RVVFSNGTLLSEAVARLV--GLASAVYQKQKPVMDLHGLALFLKQPHGLLEPKSMP 840  
 DB 730 RLIFSN--TAGSFAVOSGFVKPIGIRDY-----VVAAPYMKPLNSHKLAF----- 774  
 QY 831 PMSKSPRQWFWMLLQKQKPAOK-LTHEETITIAA-VKAFTEYALINFLNLPPTGIGE 886  
 DB 775 PUKTAAVLVLMHKISQNKRPVLRKDVDDDELAELVSVYKNSLAFEEFTNLDVVPDLSMS 834  
 QY 889 LAQFYMANLILKYCDHSUYLNLILISITIGAKRPDSSVYLHWIKKQVISAADLETQAKA 948

100 845 VARICIN(L)ACDGLDYVATL-HCLTPVLTUAREYPH-----VLSAAIATPVAV 888  
 949 LLE-----KTNLELEWTATSTHLLVAAANNORPMVVLATISIKYRGAAGNNRFQAGN-- 1003  
 889 LAFIKRGT-----ALFVOTLAKQPVATGRRLVITPVAVNYTCVGNNNVHCCNLQ 943  
 1004 WSGINGCNKVCHEFT-FHRT-----KRFELACPGCHTCVTPSSCNRETTISDY 1054  
 944 YFAGKIVDNLNLPESPSEKFTVSMLEKRRHYMT---TLEPLTPAAN--GTISTFEA 998  
 1055 KRLIVSGAMVC-----LAIVATVAVAGAAQHMADPMI-STIDPEFLARDLEHA 1106  
 999 ESKKRSVALLEKQNPENLKSITLLELRHLSGCCD-----LSSEIDVQYHGDYCMIT 1052  
 1107 POLIOTLE THWTEGALAEVAKILDEKTTAGISGETPLNLANPSCSEPHD---T 1157  
 1054 DVELETLDMIAQSGVITLH- DQCALIEDKQDADIDIQVDSDDIATASCPPEPRLPT 1109  
 1158 TSNVNISSNINSGSTVPCIKR 1179  
 1110 PS-----AGALLAKKR 1121

RESULT 6  
 AAV26945  
 ID AAV26945 standard; peptide: 45 AA.  
 AC AAV26945;  
 XX  
 XX  
 21 DEC 1999 (first entry)

DE Varicella zoster virus type 29 recombinase D35E motif.  
 XX  
 XX

Recombinogenic motif: retrovirus; integrase; invertebrate; transposase;  
 consensus: varicella; gene therapy; allergy; autoimmune disease; cancer;  
 cardiovascular disease; graft rejection; hematopoietic disorder;  
 immunosuppressive disorder; immunoproliferative disease; jaundice;  
 immunodeficiency disease; infectious disease; inflammatory disease;  
 septic shock; metabolic defect.

XX  
 XX  
 OS Varicella zoster virus.  
 XX  
 XX  
 CN US6187584-B1.  
 XX  
 XX  
 PD 28 SEP 1999.  
 XX  
 XX  
 PE 28 FEB 1997, 97US-0807332.  
 XX  
 XX  
 PR 01 MAR 1996; 96US-0012616.  
 XX  
 XX  
 PR 02 AUG 1996; 96US-0024064.  
 XX  
 XX  
 PA (NAME) NAT JEWISH MEDICAL & RES CENT.  
 XX  
 XX  
 PI Gelfand EM, Dreyfus DH;  
 XX  
 XX  
 DR WPI: 1999-561074/47.

PT peptides derived from a Herpes virus recombinogenic motif useful for the  
 PT production of vaccines and in gene therapy.  
 XX  
 XX  
 PT Claim 1: Column 47-49; 29pp; English.

XX The invention relates to novel recombinogenic motifs found in  
 XX (retroviral) integrases which have similarity to, for example,  
 XX transposase molecules. The peptides, or their homologues, have the  
 XX following identifying characteristics: (a) an initial and terminal amino  
 XX acid comprising Asp or Glu, separated by 34-39 amino acids, especially  
 XX 45 amino acids (the D35E motif); (b) a high probability ( $P < 0.05$ ) of  
 XX alignment with a consensus sequence (AAV26941), as determined by:  
 XX (1) aligning the initial and terminal residues of the homologue sequence  
 XX with those of the consensus; (2) aligning the intervening residues of  
 XX the homologue with those of the consensus by maintaining the spacing of

CC the consensus while, if necessary, altering (by the insertion of spaces  
 CC or the deletion of residues) the spacing of the homologue; and  
 CC (3) generating a distribution of 1500 random sequences of amino acids  
 CC between the initial and terminal residues and identifying the percentage  
 CC of aligned residues. The recombinogenic motifs may be useful in the  
 CC production of vaccines and in gene therapy to treat allergies.  
 CC autoimmune diseases, cancers, cardiovascular diseases, graft rejection,  
 CC hematopoietic disorders, immunosuppressive disorders,  
 CC immunoproliferative diseases, immunodeficiency diseases, infectious  
 CC diseases, inflammatory diseases, jaundice, septic shock and other  
 CC immunological, genetic or metabolic defects.

CC Sequence 35 AA:

Query Match 2.9%; Score 183; DB 20; Length 35;  
 Best Local Similarity 100.0%; Pred No. 3, 1e-10;  
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 545 DPLGNYAPYLIRKPGDQTEAKATMDIYRATLE 579  
 DB 1 DPLGNYAPYLIRKPGDQTEAKATMDIYRATLE 35

RESULT 7  
 AAB73931  
 ID AAB73931 standard; peptide: 35 AA.  
 AC AAB73931;  
 XX  
 XX  
 29 MAY-2001 (first entry)

DE Varicella zoster VZV29 protein recombinogenic sequence.  
 XX  
 XX

Varicella zoster virus; VZV29; herpesvirus; D35E; anti-viral;  
 anti-immunosuppressant; immunostimulant; cytostatic; anti-allergic; cardiac;  
 anti-inflammation; gene therapy; infection; herpes virus replication;  
 V(D)J recombination; retroviral integrase;  
 immunoglobulin class switching; allergy; autoimmune disease; cancer;  
 cardiovascular disease; graft rejection; haematopoietic disorder;  
 immunological disorder.

XX  
 XX  
 OS Varicella-zoster virus.  
 XX  
 XX  
 PN US6187584-B1.  
 XX  
 XX  
 PD 13-FEB-2001.  
 XX  
 XX  
 PE 23-JUN-1999; 99US-0338876.  
 XX  
 XX  
 PR 01-MAR-1997; 96US-0012616.  
 XX  
 XX  
 PR 02-AUG-1996; 96US-0023064.  
 XX  
 XX  
 PR 28-FEB-1997; 97US-0807332.  
 XX  
 XX  
 PA (NAME) NAT JEWISH MEDICAL & RES CENT.  
 XX  
 XX  
 PI Dreyfus DH, Gelfand EM;  
 XX  
 XX  
 DR WPI: 2001-256131/26.

PT New nucleic acid molecules encoding peptides with transposase  
 PT activities, useful as immunosuppressants, anti-viral agents, or vectors  
 PT for gene therapy of diseases, e.g. allergies, autoimmune diseases,  
 PT cancers or graft rejection.

PT Claim 1: Column 25, 31pp; English.

XX The present sequence is provided in a specification relating to a now  
 XX isolated nucleic acid molecule comprising a nucleic acid sequence  
 XX encoding a peptide consisting of one or ten amino acid sequences.  
 XX The nucleic acid molecules are useful for producing peptides  
 XX important in the regulation and function of herpes virus replication,  
 XX V(D)J recombination, retroviral integrase function or immunoglobulin

CC class switching. The isolated nucleic acid molecules and the peptides  
CC they encode are useful as immunosuppressants, anti-viral agents or  
CC vectors for gene therapy. These are particularly useful in protecting  
CC humans from a variety of diseases, e.g. allergies, autoimmune diseases,  
CC cancers, cardiovascular diseases, graft rejection, hematopoietic  
CC disorders, immunodeficiency diseases, immunoproliferative diseases,  
CC immunosuppressive disorders, infectious diseases, inflammatory diseases,  
CC jaundice, septic shock, or other immunological defects, as well as other  
CC genetic or metabolic defects.

XX Sequence 35 AA:

Query Match 2.9%; Score 183; DB 22; Length 35;  
Best Local Similarity 100.0%; Pred. No. 3, 1e-10;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 545 DPLGNAPYLILRKPGDTEAKATMODTYRATLE 579

DB 1 DPLGNAPYLILRKPGDTEAKATMODTYRATLE 35

RESULT 8

AA07562

ID AAB07562 standard: Protein: 1841 AA.

XX AAB07562:

20-OCT-2000 (first entry)

DE Protein encoded by the bleomycin (BLM) gene cluster ORF30.

XX BLM gene cluster; bleomycin gene cluster; polyketide metabolite;  
XX bleomycin, bleomycin analogue, holo carrier protein, thiazolidine;  
XX thiazoline, bithiazoline, microbial metabolite, sugar.

XX Streptomyces verticillius.

XX MO200040704-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000MO-0500445.

XX 06-JAN-1999; 9905-0115435.

XX 05-FEB-1999; 9905-0115848.

XX 05-JAN-2000; 200005-0477962.

XX (BEGC) UNIV CALIFORNIA.

XX Shen B, Du L, Sanchez C, Chen M, Edwards DJ:

XX WPI: 2000-465974/40.

XX N-PSDH; AAA58471.

XX New bleomycin gene cluster components useful for peptide and/or  
XX polyketide metabolites, especially bleomycin, production and for  
XX chemically modifying biological molecules.

XX Disclosure: Page 97-137; 162pp; English.

XX AAB07556-78 represent proteins encoded by open reading frames (ORFs)  
XX 8 to 30 of the BLM (Bleomycin) gene cluster. The proteins encoded  
XX by the gene cluster are useful for producing peptides and/or polyketide  
XX metabolites, especially bleomycin or bleomycin analogues. They are  
XX also useful for chemically modifying biological molecules to produce  
XX branched methyl groups, and for coupling amino acids and fatty  
XX acids. They may be reacted with an apo-carrier protein and coenzyme A  
XX to produce a holo-carrier protein. The BLM gene cluster or catalytic  
XX domains can be used individually or collectively to produce  
XX thiazolidine, thiazoline, bithiazoline and bithiazoline-containing  
XX microbial metabolites. The BLM gene cluster may also be used to produce  
XX sugars.

SQ Sequence 1841 AA:

Query Match 2.5%; Score 156.5; DB 21; Length 1841;  
Best Local Similarity 19.6%; Pred. No. 3, 0.0013;  
Matches 268; Conservative 133; Mismatches 470; Indels 493; Gaps 67;

OY 14 PLGVYACREDELDLEISFLARST-DSDLALPLMRNTVKTFTSSLAIV-----65

DB 461 PMLVVSARKIKALAAVAKLAAVMSAPKQIKLADILATILAGKRAPIYAAVVCCHIDIPFA 520

OY 66 ----SCAR-TTGLAGAGITLKTTHSHYPSVYVHGCKHVLPSAAPHNLIRAGNAANERP 120

DB 521 ARLLGAGKGETALPGF-----EAVFLF-PGUGTLPPEDTGKGLVADYVAFKAFHF 567

OY 121 GFSRCGGPPVDCAVE---TTGAEICTRLGLEPENILIVAVIALFKAAVWCVFLHXUC 177

DB 568 -----DACACGAPLCTDLHAALGAPADDT--RAAOPALFAVEVALAKTLMKMGV 615

OY 178 LDIIVINH--GDVIRIPLPEVOLFMDVNLVDPDENTHRSIGGFVPTPYNTGLCH 235

DB 616 RPAAMLGHSIGEVVAATLAGV-LSLPDALILVARAAQOH-TMPGKMLAVPLTPDILRP 673

OY 236 LIHD-----CV-----IAPMAVALRVNVTVAAGAAHLADENHEGAV-- 274

DB 674 LRPPEVEFSAPNAPGRGVGPPPEVAELRRLARRGVPAEELATAH-AE---HSAVERP 729

OY 275 -----LPPDITTYTFQSSSSG-----TTT-----ARGARR- 299

DB 730 LIDGFRGVLEGVRLRPRLRYV---SLITGDMAAATVTAYWLAHLRIYVRAADGILKRC 786

OY 300 NDVN-----STKSPSPSGEERLAS-----IMAADTA---LHA-EVIENTGIY- 339

DB 787 LDGVALVLEIGRKAGITGLAKRAAGCBPPIYVKCLAAVDFAASLPHAAVALMRSQCAV 846

OY 340 -----ETPTDIDK-----EIMPPIGME 356

DB 847 DMTAFHPRGPRRTYVPGYPEQVRHWDAPDESEPTDLATLAKELRTGSDPLAVDQR 906

OY 357 -GTLPRINALGTYARVAVIGAMVSPNSALYL-TEFEGSGMTEAKDGPGSPFNFYQ 414

DB 907 PGLRTGLNRLCALARD-----YLATVEASGV-----POFHAFID 943

OY 415 FAGPHLAANDPDRDGHVSSOSTGSSNTEFSVDYALILGFGAPLAPLFLYLERPDAG 474

DB 944 YLRTLAASAPAAD-DAGTTIAEITAAH-----PSRSGIVDLIRHCAGC 985

OY 475 -----AFTGGHGDALKVYTFPDSRIPCSICEKHTRPCATTHRLRQHM 520

DB 986 YPRALSTPGNALDVLTPAGSGDLRRLTG-----ETADIRATGRL 1026

OY 521 PRPG-----QATROPIGVGTMSQVSDCPICNYAPYLILKPGDOTEAKATMODT 573

DB 1027 TRLAGSLDLRLADREGRPLRVLEAGC-----AGSLTQALVLRAPG-----RLDY 1073

OY 574 YRATLELFT-----DLEQERLLDRGAPCSSEGLS-----SYVDHPT 611

DB 1074 HATDISHHFTALGKRAAKGGLDFVRAKYLDLAKDPEQDGFAGKRPVYVGLVYVHATPD 1133

OY 612 FRRILDTLRARIQOTTQOFKVLVETRDYKLRGLSATHSMALTPPYGACAPIT--- 668

DB 1134 LKRTTLGRLSLMAPDGT--LALLET-----TADDPWLLTMIGLTDGW 1173

OY 669 -NFLVKKTH-----LAVVODLALSQCCHVFYGGQVFGNPNPNO-----PQVPL 710

DB 1174 WHHTDRTRTHGLDLDAAGWMLLAGEDPATADVLVPRPGDPAALILARQPPRPAAPASV 1233

OY 711 RPRFVLFF-----NQGFI-----STESITYTISGPPSAPARPTLGGD 747

DB 1234 GKRDOVCTWCYARQWRHAAAPADPAPLTCCGLLDDQDAKAVASRIE-ALGVIVTIVCGC 1291

OY 748 APAG-----RIFDGLARVSEV--IRDIRVKNRVVSGNCTINSEARAKILVAGIASYOR 801

DB 1292 RPPGPEKYEKRELVPATRIPLAVDLWPLRLDASHKRGAAAGAGVRTQDDALJHLHLHARAFGA 1351

QY 802 QERR-----VDMHGAIGFLKQFHLFPRGMPNKSPPQWFWTLGQNPADKLT 856  
 DB 1452 LERKHARAVTVTGA-----HNYL-----GDDLA 1376  
 QY 857 DEETITL AAVKRETEY AAININLPETIGFAQFYMANLILKYTHSYLINTLT 913  
 DB 1477 DEHATVPAAAKVLPREYPIACTALIVERGLDAE-----RLADLIVERLGAAR---ETTV 1429  
 QY 914 STETARR-----PROSSVI HMLPKOVTSADILETQAKALLEETENLPHLMTAFSTHL 969  
 DB 1440 TATGRRRETTP-----VRQPLPA-----PERPAVRPGGVYL 1463  
 QY 979 VVAMNQRINNVVIGTISTEYFNA-----GNRPVTA-----INMSTLWCKKVVTFLEFDP 1022  
 DB 1464 VTCGLG-----GLHATVLCNARTTVVTHRRPDPAGMGLDAGH---PBAAYVR 1515  
 QY 1024 RPTTACPRGGTICPVTPGSSNPEETLSTQ-----VPGITVSGG--- 1062  
 DB 1516 LRSIAA-----TGATVVVRKADCTDHPAMRALADEVGAHSPVRCVHAAGVPT 1565  
 QY 1063 AAVQALATVAVRACARAOHMAFDIMLSLIDDEPLARDLEL 1105  
 DB 1566 ASMTORRKAQTALAAKLITL-----VLDEVFAHRDLDEL 1603

## RESULT 9

AAH75958  
 ID AAH75958 standard; Protein; 3210 AA.

AC AAH75958;

XX 29 MAY 2001 (first entry)

DE Mycelia sterilia cyclic depsipeptide synthase.

XX Mycelia sterilia: cyclic depsipeptide synthase; antihelminthic;

KM PF1022.

XX Mycelia sterilia.

XX W0200118179-AL.

XX 15 MAR 2001.

XX 07 SEP 2000; 2000MO JP06103.

XX 07 SEP 1999; 99JP-0254040.

XX 06 APR-2000; 2000AP 0104291.

XX (MED) MELJI SEIKA KAISHA LTD.

XX Mitsu N. Okakura K, Miyamoto K, Watanabe M, Yanai K, Yasutake T,

XX Althara S, Furumura T, Kleinkauf H, Murakami T,

XX WPL: 2001-265979/27.

XX N P50B; AAF79702.

XX Novel cyclic depsipeptide synthase and gene encoding it for efficient

XX production of antihelminthic substance PF1022

XX claim 1; page 59-60; 92pp; Japanese.

XX The present sequence is a cyclic depsipeptide synthase from Mycelia

XX sterilia. The present protein, or a protein containing an addition,

XX deletion and/or substitution of one or more amino acid residues is useful

XX for the efficient production of the antihelminthic PF1022 (cyclic(D-lactyl-

XX L-N-methylleucyl)-D-4-phenyllactyl-L-N-methylleucyl-D-lactyl-L-N-

XX methylleucyl)-D-4-phenyllactyl-L-N-methylleucyl)).

XX Sequence 3210 AA;

XX Query Match 2.0%; Score 121.5; DB 25; Length 3210;

Best Local Similarity 18.7%; Pred. No. 0.56;  
 Matches 11; Conservative 79; Mismatches 227; Indels 177; Gaps 24.

QY 298 RRDVNSTKSPSSGGERRLASTMAADTALHAHVFENTGIYERTTPDIKRWMPICME- 356  
 DB 2550 KRAVIAKKSQAAPA-----KPKVVDPEVNLNLAIT-----CEEFVILMEV 2591  
 QY 357 GTLPRIINALSYTARVAVGAVGAVFSPNSALYLTEVEDSGMTEAKDGPPSPTRFYGA 416  
 DB 2592 CVNDHPVQLGCHSLATKLVARLSRLNCRVSVRDVTDQ 2630  
 QY 417 GPHLANPQTRDGHVLS-----QSTGSSNTEFSVDYALICGFAPLLARLL 465  
 DB 2631 PVSLCLAVTLPGTITLFINLIFATISGYWEQTMSAFETISLDEAVLKEPVLVLG 2686  
 QY 466 FYLRKCIAGAPFGCHG-----DALKVYGTIDSEIPCSLCKKHTRPVCAHT 511  
 DB 2687 VEVSTDSFEDLGSHSMATKLAAPISERLDVPSYIDIDHSVFLMAK 2747  
 QY 512 TVHRLKRMRFQATROPUGVEFTMNSGYSDCPDGNVAPYLL-----PKPSD 561  
 DB 2738 -----IRLTQAKGHEATN---GV-----QIANDAPQOLISVHDEPIYVORLIAP 2778  
 QY 562 QTEAAKATMODTYRAT-LERLF-----IDLEGRLLDRGAPSS-EG 602  
 DB 2779 QLCSSPETILDVYPATQMGVHFLNPYTGKPSPTPHIDFPD-----ADCAILMRAC 2832  
 QY 603 SSYIVDHPFRRLIDPIRAPIEOTTOFMKVIVETRYKIRKGISVATHS----- 652  
 DB 2833 ASLAKHFDITRYVLEAKGELYOVVILKHVDPIEM--LQTEENINSATRSFLVDADKPT 2890  
 QY 653 -----MALTFDPSGACPTTNFLVKRTHLAVVDLALS-----QCHGVFGQVY-EG 699  
 DB 2891 RLGPILRIALILKPGS-----TLRVILRLSHALVGLSILHLIASHLILFGGSLPP 2944  
 QY 700 RNRNDFQVPL--KRRVLDLPNGSF--TKSITYTLSEGPSAANPTLGDAPAGRTTGD 757  
 DB 2945 PKFAGYMOHVAASSRRECYDWR--SVLRDSSMTYIKGNNTIPPEPPVQUSTPSGHHASK 3003  
 QY 758 IARVSVEVIRDIRKRVNVSNGCT-----NISFARARIV---GLASAVO 800  
 DB 3004 VVITPQANTDSRITRATITTTACALMLAKEDNSDVVPGRTVSGRGPLAHO 3057

## RESULT 10

AAV00206  
 ID AAV00206 standard; Protein; 1074 AA.

AC AAV00206.

XX 20-APR-1999 (first entry)

XX Enterococcus faecalis protein IP102.

XX Enterococcus faecalis, infection; vaccine; immune response; diagnosis;

XX detection; attenuation; antigenic.

XX Enterococcus faecalis.

XX W09850554-A2.

XX 12-NOV-1998.

XX 04-MAY-1998; 98WO-US08959.

XX 14-NOV-1997; 97US-0066009.

XX 06-MAY-1997; 97US-0044031.

XX 16-MAY 1997; 97US-0046655.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Bailey C, Choi GH, Bromocky J A, Kunsch CA;



used for detecting Enterococcus antibodies in a sample. The nucleotide sequences can be used for detecting Enterococcus nucleic acids. Products from the present invention can also be used for screening compounds to identify agonists and antagonists of E. faecalis protein activity.

XX Sequence 1074 AA:

Query Match 2.0% Score 123; DB 20; Length 1074;

Best local similarity 14.8%; prot No. 0.092; Matches 206; Conservative 150; Mismatches 415; Indels 352; Gaps 52;

```

UY 250 RYRNTAVARGAHLAFDENHGAVALPDITTYFGSSSGTTTARAPRNVTSTKPS 309
DB 62 RTSLVAEYNGAKUVECC-IEKVSILPTEVTHGY-----QKNPL 104
UY 410 ISGGERKSLAM AADALAAEVLFGVYEE-----TPPDIKEMPMI 353
DB 105 RSMRKAALVSLWEKAGTIDITNVAOKRMIEVNGYKLSIKRUGASVDIK----- 158
UY 654 GMSGTLPRINAGSYAR -----VAGVIG-AMVSPNSALYIEVEDSGMTEAKIGCP 405
DB 159 STECKINK ALREYCKKSHNTIKTIAGSTLILKNEMLSPEDVYVONFA----- 211
UY 406 GTSNRYFAPRPHLAANVQTDKGVLSSTGSSNTERSVDYLLIGFGAPLLARLL 465
DB 212 - NIDYVIGNOVLITP-----NSNSKQTLILKKSAGCTTP-VAAYKK 251
UY 466 FYLEPDACAFPTGCHZMAIKYVGTFTSRLPSSICKETPRVCAHTYHRLHGMRPFO 525
DB 252 ARIQYMAVALDKPNTYALIKINVEIKGS-LKIKIDKESNDIPEVPHL----- 304
UY 526 A-----TPQPIGVCTIMNSOYSDIMHICNYA-----PYLILKKPDIDTEAKATM 570
DB 305 ALPSKIVITDKIET-----SLIAGPHGCKVITTEKSVDPYMIIDITPMALIKAGETI 358
UY 571 QUTYATLTERLEFDLPHEDLRCAGCSSEGIS-----SVIVDHEPT-----FRILDTLRA 621
DB 459 SMTSKMMKOKGILLKIG-VEGTGILMNDNTSLACNTFAIKRDSAGELVOHITTDDEG 417
UY 622 KLEQTTOTPMKVIYELIRYKIRKQISLASHMALTFDP-----VSGACPTINPLVKKRI 675
DB 418 KAF-TPKELANALDELITYYTE-TRSSNGCVNTEKPTKVELKRVANQVALVTSVVRKQ 473
UY 676 HIAVVDIAL-----SOCHVYFYGOV-----KCHNPNQOQVILKRRFVIL 717
DB 474 NQHTIGETTLIKEDKQDNGESOCKAPFKGAEVLTETAKDQAVKMSAPK-----TEL 526
UY 718 FNGQFISTKSTIVTISE-GPVASPNPTIGQ-----DAPAGTFDDDLARVSE----- 764
DB 522 VGTGASDDEVTILADENKQAVAKHAIINRYMQUETKAPRGYTLDETKRYPVSIKKYDNNF 586
UY 765 - VIRDIRVNNRV-----ESQCNCTNLSSEA-----KAILVGLASAYOQEKRR 805
DB 587 KNAVITRDTAAEYIRGEDEFKFAAGMAYAEETGENDLSKVSLEGTIXETIAGEDIA 646
UY 806 VIMELHGALEP-----LAKOFHGLIFPGM-----PPNSKS-----PNP 838
DB 647 TACNPGUQDQKRYKFNENLPGVYLLLEHFA-----PRGPHKTIPLLEKSIKPKKKDYAAS 703
UY 849 GWEHTLLQHNQ -MPADKLTHERITTLIAVKTETEYAAINFLNP--PTGIGE 888
DB 704 EYVFTLIEKQKQUPINMVLVPEKILNNH-----EVSINIKMLIYDDEKRDSTHS 754
UY 889 LAQFPMANILIKYTRHSOYLINTITSITGAKRPRPSSVLHMKD---VISAAILEHQ 945
DB 755 LATMKMKNKILNTLQTE-LVQKL-----RYNLLEIKEDWYVQAQALDEVA- 799
UY 946 AKALEKTENIDELMIAFIST-----HIVRAANQORIMVIGISISKYHGA 992
DB 800 TKAAGPKPKKAPVYIAGTATATANKKRTGTAKLIKILAKO-----VLKDSIVLFPNV 853
UY 994 ANNNVAPQAGNMSLNCKKNV-PLFTFDRTRFIIACPRKAGTICPVTPSSGNRETLSD 1052

```

```

DB 854 YENKVAFEANE-----PVA-----KIANLNN 875
UY 1053 UYRQIIVSGAMVUAIYAVRAVAKR-----AQHAFIDWISLIDDEPLAKIL-- 1102
DB 876 QAO-----TVNCTIERHVSLOTKAHLEDGSQTFTHSDVMDPDDVSVTHVLD 923
UY 1103 ---EELHDIITQLETPWYVGALFAVKILDE-----KITAD---GETPINLAFNFD 1149
DB 924 GSKAFHTTLYALLDQTKNPIKRSKTHHVNKERTKIVIAKRYNGKVFID34FIR- 982
UY 1150 SCEPSHDITSNV-----LNISSNIGSTVPGLRKPPEDDE 1185
DB 983 -TEINVEKDGNVNKKIINEDLKEKSQTLTPREVPTIPSTPOPE 1024

```

# RESULT 12

ABP43407  
ID ABP43407 standard; Protein; 1074 AA.

AC ABP43407;

XX 05-AUG 2002 (first entry)

DE E faecalis EF094 protein.

KW Enterococcus; vaccine; gastrointestinal disease; diagnosis; antibiotic.

XX Enterococcus faecalis.

PN US2002045737-A1.

XX 18-APR-2002.

XX 04-MAY-1998; 98US-0071035.

XX 04-MAY-1998; 98US-0071035.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Choi GH, Bailey C, Hromockyj A, Kunsch CA;

DP WPI: 2002-425450/45.

XX N-PSDB: ABN98163.

XX New genes and polypeptides from Enterococcus faecalis, useful as

PT vaccines for preventing, treating or attenuating an infection caused by

PT a member of the Enterococcus genus in an animal, particularly E.

XX faecalis

PS Claim 9; Page 163; 255pp; English.

XX The present invention provides the protein and coding sequences of a

XX number of polypeptides from Enterococcus faecalis. The proteins can be

XX used as vaccines for preventing or attenuating an infection caused by a

XX member of the Enterococcus genus in an animal, particularly E. faecalis.

XX The polynucleotide is also useful for preventing or treating E. faecalis

XX infection. The present sequence is a protein of the invention.

XX Sequence 1074 AA:

Query Match 2.0% Score 123; DB 23; Length 1074;

Best local similarity 18.3%; prot No. 0.092; Matches 206; Conservative 150; Mismatches 415; Indels 352; Gaps 52;

```

UY 250 RYRNTAVARGAHLAFDENHGAVALPDITTYFGSSSGTTTARAPRNVTSTKPS 309
DB 62 RTSLVAEYNGAKUVECC-IEKVSILPTEVTHGY-----QKNPL 104
UY 410 ISGGERKSLAM AADALAAEVLFGVYEE-----TPPDIKEMPMI 353
DB 105 RSMRKAALVSLWEKAGTIDITNVAOKRMIEVNGYKLSIKRUGASVDIK----- 158

```







DB 575 ---SCDVGRC-----TFSHRRAMLVDOOTINMFTPLNSMEGCGN----- 610

QY 1062 GAMVLAIVATVAVGAKVAMAPD-----WLSLIDDEFLARDELIDLIQFLERI 1115

DB 611 GAKLELAHSISEAVLFEYGMALDNPNNLTIWEAOFSD--FAMAVAILIDFTIVSGET 668

QY 1116 PWTIVGAL-----EAVKILAEKFTACDGET 1140

DB 669 KWMESNALVMLPLPHGYDGAASEHSRIERFLQCDSEKETSADGDS 714

RESULT 15

AAV53665

AAV53665 standard; protein; 2387 AA.

AAV53665:

22-FEB-2000 (first entry)

Mechanical stress induced protein 608 amino acid sequence.

Mechanical stress; gene library; protein 608; osteoporosis; bone density; bone development.

Rattus sp.

Key Location/Qualifiers

Misc-difference 7 /note- "encoded by CAT"

Misc-difference 363 /note- "encoded by GGA"

Misc-difference 1453 /note- "encoded by GGA"

Misc-difference 1453 /note- "encoded by NTT"

MO9960164-A1.

25-NOV-1999.

14-MAY-1999; 99MO-US11066.

15-MAY-1998; 98US-0085673.

(QJAK-) QJAK H10TECH INC.

Elinal P, Mor O, Skallier R, Feinstein E, Faerman A:

WPI: 2000-054404/04

N-PSDB: AAV36322.

Identification of stress induced genes for determining risk and preventing, treating or controlling osteoporosis

Claim 32, Fig 5A-Z, 308pp, English.

The present sequence represents protein 608, which was identified using the method of the invention after subjecting rat osteoblasts to mechanical stress. Expression of the 608 gene was found to be upregulated by about 3-fold in cells subjected to mechanical strain. The specification describes a method for the identification of genes responsive to a specific mechanical stress. The method comprises applying the mechanical stress to an organism (tissue or cells comprising bone cells), isolating the specific cellular fractions and extracting RNA from them, and differentially analysing the RNA in comparison with control samples. The method is used to identify genes whose expression is responsive to a specific stress. The identified genes are employed in determining risk associated with a physiological or disease state. The risk determination methods are used for testing a medicament for gene therapy. These medicaments, or genes identified by the method of the invention, are used for treating, preventing or controlling a physiological or disease state (especially osteoporosis or bone density or other factors causing or contributing to osteoporosis or its symptoms or other conditions involved in mechanical stress or its lack. The methods can also be used for advancing research or studies in bone

CC development.

XX XX

SV Sequence 2387 AA;

Query Match 1.9%; Score 122.5; DB 21; Length 2387,

Best Local Similarity 18.6%; Pred. No. 0.42;

Matches 145; Conservative 114; Mismatches 286; Indels 233; Gaps 40;

QY 416 AGPHLAANPOTD---RDGVLSQSSTGSSN-----TEFSVDYLAALIG 455

DB 64 SGAFLOCTKPTIDPSLSKSLVTQEDNGSASTSPQDEIEPFGSISLMTLSGNKADWVS 123

QY 456 FGAFLIARILFYLERCDAAGFTGHDALKYVGFPESEIPCSLCRKHTRPV----CAH 511

DB 124 IQKP-----SRSPAFTEENDYIM--LNASTNLVGVSDVNHQOPVQALLAYS 172

QY 512 TVHRLRQMP-----KEGQATROPVIGVGTNNSVSDCDPLGANTAPYILKPGD 561

DB 173 DSPILIERKPOLTEIPLSSSKYKQVALRPEDLFTSLIAVVK-ADPWPQOEKIVQINRT 231

QY 562 QTEAAKATMODYRATLEKFLIDLEGERL-----TLK-----GAPGSEJ 601

DB 434 ATTLSTLQIUFSTDAIALPRAEMPAEFLKWTMI--AMNNKLEFIVLVATIALSLP-K 291

QY 602 ----LSSVIVDHPTR-----RILDTLAKILR-OTTTPMKVL---VETRD----- 639

DB 292 DSPHLEWILLACSSKVPARYVSEF3PILIDKN-KLELMAMSFTFASLYHGISSTADADV 351

QY 640 --YKIR--EILSEATSMALTFDPYSGA-----CPTNPLVKRTHLAVYODLALSGCHV 691

DB 352 LTVRTIVVEYGCSTHDSGOHTVGTETLDPLCLTG-----VIDASIS--W 398

QY 692 FYGOQVEGRNFRNQFVLRREFDEFGGFISTRITVTLSEGPS--APNPT----- 743

DB 399 LPGNTEVFSQPSRDR-----QILNCTLRILQVT--PKDGHVQCVANPSGAEFSS 447

QY 744 -----LGDAPRGRITPDGLARYSVEVIRIRKYNVAVSGNCTINSEAKAR 791

DB 448 FKVSVOKKQGRWEHDEAGSGSLGE--PNSSVSLKQASLK-----LSASALTSGEAKQ 501

QY 792 LVGLASAVYUQERKVDMLNGLG--FLIKUFHGLLEFGMPNSKSPNUPFWTLQ---R 847

DB 502 ----VSGVHRKKNHRLIHRRGOSTLRPRP--HRHQPLISARIPDQWAAALFEKAK 555

QY 848 NQMPADKLTHEITT-----IAVKKFTEEYATLFTNLPPCLSELALQFYANILK 900

DB 556 NSVP-----KKQENTYKPVPLAVPELDELDEKDAQMI--FP----- 592

QY 901 YCHSGVLIHLNLSLTGAKRPKUPSSVLMIRKDVISAADIEQAALIEKTNLEIM 960

DB 593 --DEEFVVLTKASGVGKSPADSGVNNGFMTISAGIEVSVNPUTLQ--SEHLPDEK 649

QY 961 TTAFTSTHLVRAM-----NQRPMVVLGISIKYHIAAANNKPAQANMSG-- 1006

DB 650 LFSVTNQTAVTKMSNPISAKIEDTQNP111FP-SVAFIPDQA-----QACRASSQS 702

QY 1007 ---LNGSKNCPLETFTRTPRTIACPRGGFI-----CPVTG---PSSGN 1045

DB 703 AHPVTGN---MATYGHNTYSSFTSKASTVLQPLNFTESYGVQITITGVSRPSSD 756

Search completed, March 28, 2003, 13:35.35

Job time: 96 secs



GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using SW model

Run on: March 28, 2003, 13:33:59 ; Search time 31 Seconds  
(without alignments)  
1141.799 Million cell updates/sec

Title: US-09-769-699-2  
Perfect score: 6294  
Sequence: 1 MENTOKTYVTGTPGLGVYA.....DELFDLSGIFIKHGNITMEM 1203

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Minimum number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents-AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCUTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfiletest.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3054	48.5	1452	4	US-09-127-227-2
2	183	2.9	35	2	US-08-807-332B-9
3	183	2.9	35	4	US-09-338-876-9
4	131	2.1	757	4	US-09-413-814-84
5	123	2.0	1074	4	US-09-071-035-358
6	123	2.0	1074	4	US-09-071-035-394
7	111	1.8	5215	4	US-07-642-734C-4
8	106.5	1.7	3567	3	US-08-439-009A-4
9	106.5	1.7	3567	3	US-07-731-157A-7
10	106	1.7	774	1	US-08-541-780-7
11	106	1.7	774	2	US-08-633-760-46
12	105	1.7	774	1	US-08-633-760-48
13	105	1.7	774	4	US-08-931-608A-5
14	104.5	1.7	635	4	US-08-019-870-5
15	104.5	1.7	774	1	US-08-019-870-3
16	103	1.6	774	1	US-08-633-760-44
17	103	1.6	774	1	US-08-070-165F-6
18	102	1.6	729	2	US-08-885-418-6
19	102	1.6	729	2	US-07-747-801A-3
20	102	1.6	774	1	US-07-945-312-3
21	102	1.6	774	1	US-08-633-760-50
22	101.5	1.6	773	1	US-08-019-870-1
23	101.5	1.6	773	1	US-08-019-870-6
24	101.5	1.6	773	1	US-08-019-870-8
25	101.5	1.6	774	1	US-08-019-870-11
26	101.5	1.6	774	1	US-08-633-760-52
27	101.5	1.6	774	1	US-08-633-760-52

28	101	1.6	1024	4	US-09-562-737-4R	Sequence 48, Appl
29	101	1.6	7257	3	US-09-335-409-5	Sequence 5, Appl
30	101	1.6	7257	4	US-09-568-102-5	Sequence 5, Appl
31	101	1.6	7257	4	US-09-567-969-5	Sequence 5, Appl
32	101	1.6	7257	4	US-09-568-480-5	Sequence 5, Appl
33	101	1.6	7257	4	US-09-568-486-5	Sequence 5, Appl
34	101	1.6	7257	4	US-09-568-472-5	Sequence 5, Appl
35	101	1.6	7257	4	US-09-567-899-5	Sequence 3, Appl
36	100.5	1.6	1231	5	PCT-US95-07744A-3	Sequence 3, Appl
37	100.5	1.6	1321	5	PCT-US95-07744A-3	Sequence 3, Appl
38	100.5	1.6	2616	6	5206163-3	Sequence 5, Appl
39	99	1.6	915	1	US-08-328-322-5	Sequence 3, Appl
40	98.5	1.6	2152	4	US-09-336-987A-3	Sequence 3, Appl
41	98.5	1.6	2152	4	US-09-370-700-3	Sequence 13, Appl
42	98	1.6	821	2	US-08-451-822A-13	Sequence 13, Appl
43	98	1.6	821	4	US-08-323-430-13	Sequence 13, Appl
44	97.5	1.5	769	1	US-08-471-570-8	Sequence 8, Appl
45	97.5	1.5	1024	4	US-09-562-737-50	Sequence 50, Appl

## ALIGNMENTS

RESULT 1  
US-09-127-227-2  
Sequence 2, Application US/09127227  
Patent No. 6399354  
GENERAL INFORMATION:  
APPLICANT: David M. Knipe  
APPLICANT: Travis J. Taylor  
APPLICANT: Elizabeth McNamee  
TITLE OF INVENTION: Replication-Competent Virus Expressing A  
FILE REFERENCE: H098-05  
CURRENT APPLICATION NUMBER: US/09/127,227  
CURRENT FILING DATE: 1998-07-31  
NUMBER OF SEQ. ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ. ID NO. 2  
LENGTH: 1452  
TYPE: PRT  
ORGANISM: herpesvirus  
US-09-127-227-2

Query Match 48.5% Score 3054; DB 4; Length 1452;  
Best Local Similarity 49.8%; Pred. No. 1.8e-304;  
Matches 609; Conservative 200; Mismatches 360; Indels 46; Gaps 17;

QY	1	MENTOKTYT---VPTGPGYVY--ACRYEDLDLEISFLARSTDSIALPLMKRLTVE	55
DB	1	MENTOKTYT---VPTGPGYVY--ACRYEDLDLEISFLARSTDSIALPLMKRLTVE	55
QY	56	KTFSSLAIVSCARTTGACAGCITLKTSHFVSVFPGCKHVLPSAPNLTTPACNA	115
DB	56	KTFSSLAIVSCARTTGACAGCITLKTSHFVSVFPGCKHVLPSAPNLTTPACNA	115
QY	59	SGFEENVAIVVSSRTTGGTAVSLKLTPSHSSVVFHGGRIHDPSTQAPNLTFLCER	118
DB	59	SGFEENVAIVVSSRTTGGTAVSLKLTPSHSSVVFHGGRIHDPSTQAPNLTFLCER	118
QY	116	ARERGFSCQCPVDDGAVETTGAEICTRLGEPNTILYLVFALFKEAVFMCVPLAY	175
DB	116	ARERGFSCQCPVDDGAVETTGAEICTRLGEPNTILYLVFALFKEAVFMCVPLAY	175
QY	119	ARRHGFISYTRPDLDKHEITGECALGIDPPRALILYLVTEGFEACVACINNTFHL	178
DB	119	ARRHGFISYTRPDLDKHEITGECALGIDPPRALILYLVTEGFEACVACINNTFHL	178
QY	176	GGIDLVTHNGDVIRLPLEFVOLFMEDVNRVLVDFPNTHHKSIGSEGYFFPYFTIGLCH	235
DB	176	GGIDLVTHNGDVIRLPLEFVOLFMEDVNRVLVDFPNTHHKSIGSEGYFFPYFTIGLCH	235
QY	179	CGSDKVTTCGACVHRIPVPLQLFMDPSPVLAEPFNHRSIGNFPYPLDPFPRPILNR	238
DB	179	CGSDKVTTCGACVHRIPVPLQLFMDPSPVLAEPFNHRSIGNFPYPLDPFPRPILNR	238
QY	236	LHDCVIAIMAAVALFVRNVTAVARGAHLAPFNNHGAVALPDIYTFYQSSSSCTTAR	295
DB	236	LHDCVIAIMAAVALFVRNVTAVARGAHLAPFNNHGAVALPDIYTFYQSSSSCTTAR	295
QY	239	LLEAVVGAALAAKLRNDATAARAAHLAPDENEGALAPDITFAFENSG--KIPK	296
DB	239	LLEAVVGAALAAKLRNDATAARAAHLAPDENEGALAPDITFAFENSG--KIPK	296
QY	296	GARRNDVNSTKSPSGGFPFRIASTMAADTALHAFLVINTGIVERTPDIKKMMFICM	355
DB	296	GARRNDVNSTKSPSGGFPFRIASTMAADTALHAFLVINTGIVERTPDIKKMMFICM	355
QY	297	CGR-----DCCGCKPAGCFEORLASVMAGDAALALEISVSMVFPNPPPDJISAMIPCGCQ	351
DB	297	CGR-----DCCGCKPAGCFEORLASVMAGDAALALEISVSMVFPNPPPDJISAMIPCGCQ	351
QY	356	EGTLPLNLGLSTAVAVAGVAMVSPNSALYLTVEEDSGMTAKKDGSPGSPFRFQF	415
DB	356	EGTLPLNLGLSTAVAVAGVAMVSPNSALYLTVEEDSGMTAKKDGSPGSPFRFQF	415

```

100 452 D1AAANAVGATLARAAGVCAWVSTNSALHLEVDIAGPADPKHCK-PSYRRTIV 410
101 416 ACPIILANPOTODHIVL-----SSQSTGSSNTEESVYALITGFPAPLLALLFYL 468
102 411 PCHIVANAVQVREKHNVAQVPRKRPAPVVOGQV-EPAGHILAMIDQFSAILLAKMLFYL 469
103 469 ERTDAGAPITGHC DAKVYVCTIPISEIIPCISCEKHTKRVCAHTVH QIKOMRRCQAT 527
104 470 EKKDKGVIVGRQEMQVFRVVDNSQTDVPCNCTFTFRHACVHTTLMRLARHPKVASNA 529
105 528 KQPIVGVGIMNSYSKRDQLANAPVLLIKRQVQTEFAKATIMQDITKATLEPLDQ 587
106 540 KATIGVPGIMNSYSKRDQLANAPVLLIKRQVQTEFAKATIMQDITKATLEPLDQ 588
107 588 EPIILGAPESSELSVIVIEFTFRIILDELARLEOTTQPMKLVETROKIREGELS 647
108 589 EGYVQAVTAMCHLEITINHALITVANNVQVVDKREKRLMNVKRNKPKDGLG 648
109 648 EATSKALTFIDYSCAPVPIINLYKRTIHAVODIALSQCHECVYQGVQECNPNQFO 707
110 649 EANHMSLIDPYATGCPCLDILGRSNLAVYQDLALSQCHECVYQGVQECNPNQFO 708
111 708 PVIIRHFPVLPNGPSTSTVTITSEG-PVSAPNPTIGQDAPACRTPQDILARSVVVI 766
112 709 PVIIRHFPVLPNGPSTSTVTITSEG-PVSAPNPTIGQDAPACRTPQDILARSVVVI 768
113 767 KOLIKKNRVESGNTNLSAARARVLAQAVQPEKRVDMHGLAFELKQFHLLEP 826
114 769 KKLKAKSVILFANASANASAKAVASVQYQKPKRVDDILGLQHLKQFHAIIP 828
115 827 KQMPNRSKSNQVWVTLQORQVADKTHRETTIAVKKRTERYAALNPNPTG 886
116 829 NKRPSSNQPNQVWVTLQORQVADKTHRETTIAVKKRTERYAALNPNPTG 888
117 887 CEIAPYMANLHKYQDHSOYLINLTSTIGARPRDSSVILHVRKQVISAADLETQA 946
118 889 SELAMYMANQULPRQDHSYFINTITAIACSRHPSVQAAAM--SAQCGAGIACAA 945
119 947 KALLKTEMLPELTATATSTILVRAMNORPVVLTGISYHGAQGNRRVQAGNMSG 1006
120 948 KALIMAVIAHAPAMISMAASCLIKRVMAKRVVIGLSTSKYQAGMDRVQAGNMS 1005
121 1007 INGCNVCPLEFDEKRRFLIACPGCGFICVTCSSCKRFTTSNOVCGITVSCGAMQ 1066
122 1006 LMCKNATCELTIFORTKRVFLVAPRAGVTAASNLGSAHSSLECELGITSEGAAYA 1065
123 1067 LAIVTVRAVAGARADHMAIDMISTIDDELARLEPLDQ 1071 PPTVFCAL-- 1123
124 1066 NSVPAVATKSHDRTIGQVPIVWALTEDEYSSEPMELTARALKNGCNSWDALFVA 1125
125 1124 EAVKILDEKTTVAAGPPTNIAFNPD--STPSHITTSNVLINISGNSGSTRVPLKRP 1180
126 1125 HFAKALVLSQANAGE--VFNFQDQGR-----DQNAITPQDQAPGAPAFAGKKA 1173
127 1181 PEDDELFDLSIPEKKNITMIM 1203
128 1174 EFGDEFG EFGDEFGDGLTDM 1195

```

```

RESULT 2
US 08 807 4428 9
Sequence 9, Applied for US/088074428
Patent No. 5959074
GENERAL INFORMATION:
APPLICANT: Dreyfus, David H.
ATTORNEY/AGENT INFORMATION:
GELLAND, ERWIN W.
TITLE OF INVENTION: PRODUCTS AND PROCESSES FOR REGULATION OF
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSER: Sheridan Ross
STREET: 1700 Lincoln St., Suite 3500

```

```

? CITY: Denver
? STATE: CO
? COUNTRY: U.S.A.
? ZIP: 80203
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/807,332B
? FILING DATE: 28-FEB-1997
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Kovarik, Joseph E.
? REGISTRATION NUMBER: 33,005
? REFERENCE/DOCKET NUMBER: 2879-39
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 303/863-9700
? TELEFAX: 303/863-0223
? INFORMATION FOR SEQ ID NO: 9:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 35 amino acids
? TYPE: amino acid
? STRANDEDNESS:
? TOPOLOGY: linear
? MOLECULE TYPE: peptide
? US-08-807-332B-9
?
? Query Match 2.9%; Score 183; DB 2; Length 35;
? Best local Similarity 100.0%; Pred. No. 1,1e-11;
? Matches 35, Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

DB 1 DPLGNVAPVLLIKRPGDQTEAAKATIMQDITKATLE 35
US-09-338-876-9
Sequence 9, Application US/09338876
Patent No. 6187584
GENERAL INFORMATION:
APPLICANT: Dreyfus, David H.
ATTORNEY/AGENT INFORMATION:
GELLAND, ERWIN W.
TITLE OF INVENTION: PRODUCTS AND PROCESSES FOR REGULATION OF
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSER: Sheridan Ross
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: CO
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/338,876
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/807,332
FILING DATE: 28-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Kovarik, Joseph E.
REGISTRATION NUMBER: 33,005
REFERENCE/DOCKET NUMBER: 2879-39
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/863-9700

```

TELEFAX: 303/863-0223  
 INFORMATION FOR SEQ ID NO: 9:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 35 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-09-338-876-9

Query Match 2.9%; Score 183; DB 4; Length 35;  
 Best Local Similarity 100.0%, Pred. No. 11e-11;  
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 545 DPLSNVAPYLIRKRPDTEAKAIMODYRATLE 579  
 DB 1 DPLSNVAPYLIRKRPDTEAKAIMODYRATLE 35

SEQUENCE 4  
 -09-413-814-84  
 Sequence 84: Application US/09413814

GENERAL INFORMATION:  
 APPLICANT: Gesellschaft fuer Biotechnologische Forschung MBH  
 APPLICANT: Bristol-Myers Squibb, Co.  
 APPLICANT: Bayer, Stefan  
 APPLICANT: Bloecker, Helmut  
 APPLICANT: Brandt, Petra  
 APPLICANT: Cino, Paul M  
 APPLICANT: Dougherty, Brian A  
 APPLICANT: Goldberg, Steven L  
 APPLICANT: Helle, Gerhard  
 APPLICANT: Mueller, Joachim  
 APPLICANT: Reichenbach, Hans  
 TITLE OF INVENTION: heteropolypeptide compounds  
 TITLE OF INVENTION: heteropolypeptide compounds  
 FILE REFERENCE: PCT/US 99/23535  
 CURRENT APPLICATION NUMBER: US/09/413,814  
 CURRENT FILING DATE: 1999-10-07  
 EARLIER APPLICATION NUMBER: DE 198 46 493.2  
 EARLIER FILING DATE: 1998-10-09  
 NUMBER OF SEQ ID NOS: 107  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 84  
 LENGTH: 757  
 TYPE: PRT  
 ORGANISM: Sorangium cellulosum  
 -09-413-814-84

Query Match 2.1%; Score 131; DB 4; Length 757;  
 Best Local Similarity 21.1%, Pred. No. 0.00079;  
 Matches 146; Conservative 84; Mismatches 251; Indels 212; Gaps 37.

QY 54 VKITFTSSIAVVSQAT-----IGLACGTLTKLITSHFPSPVVFHGGHVLVSSAPNI 109  
 DB 129 LERTLPRAVSIADAAAEALITSSAVAGIVAS-----LPASAAA-- 167  
 QY 110 TRACNAAREPFESRQGPVDAVETTSAGICTRLGLEPENTILYLVATLFEKAVEM 169  
 DB 168 -----AAQIRLWIAIDQS--PGPIHGRCA-----ALKRPSV----- 198  
 QY 170 NVFLHY-----GGLDIVINH-DEVTIRPLEPVGLFMPVQNRIVPPFNTIHRSIRGPVY 244  
 DB 199 AFLDYTSSTGDEPKVMILHNL-----LHNSRLTAHGFDLTSPDPV-----GVTW 244  
 QY 225 PVPVYNTGIC-----HIHDCVIAPMVAVALR-----VRNTAV-----ANGAHLAPV-- 267  
 DB 245 LPPYIDMGILGILVALYRIRIVALLSPISFQRPFRKRAVASALCASVSGGFENFAYDLG 304  
 QY 268 -----ENHICAVLPDITTYFQSSSSSGITTAGARRDNVNSTSKSPSSGFEKRLASIMA 323  
 DB 305 VKSSHFPAALD-----LNSWVAFPGADPRVADTLDPFARARAVASGPRPR--AFYP 355

QY 324 ADTALHAAYLEFNIGIYEETPTDIKEMPMETGMHCTLPRIALGSTARYAGVIGAMP-- 381  
 DB 356 CYGLAEATLIVSGGAREAPV-----TARLAPEEVELEIRAVASAAE--GARFVG 403  
 QY 382 -----SNVALYLFVEVDJGMEFAKIDGCPSPFNKRYQFAGPLA-----ANVDTIDGHV 432  
 DB 404 SKKALLPKA---VAIVDPAG-----NELGCG-EIGLHW-VSGSVAAVQWGRPP----- 447  
 QY 433 LSSQSTGSSNTERSVYDYLALLI--FAPILARILFLF--TA--FT--aHSLKLVYGT-- 490  
 DB 448 -----ETEAIFGAILASSAAR-----YLTGIDDELKQDE-----LFVDRSK 486  
 QY 491 -----PDSEIPCSICEKH--TRPVC--ANTVHRIKQM-----PPRGQATKG 529  
 DB 487 DLTILGRNFFPDITKTESSHPAARPGCSAASFVEHGEFPLAVVCEYDPRVADPRE 546  
 QY 530 PIGVPGTMSQVSIQDPLSNVAPYLIRKRGVIFAKAIMODYRATLEKLR--TDLPOER 589  
 DB 547 IVAAREAVTAEHQ---LVAHAVALIAPCALPKTSSGKVRHRCRAFLR---DALGR 598  
 QY 590 -----LDRGAPPSSEGLSVYVDHPPTFRPLTLTPAPTEGTTGFMKVLVETDQKT 642  
 DB 599 HVAFAPELLDDASPPD---APPETEPPSGRSLILALRS---TLAKALRLDAGQIDDAL 651  
 QY 643 ---KEGL-SHATHSMALTFUPYSGAPCIHNF 671  
 DB 652 PISRFGLSLAAYELQHAFOVIRGRAIPILSIL 684

RESULT 5  
 US-09-071-035-358  
 Sequence 358: Application US/09071035  
 Patent No. 6448043  
 GENERAL INFORMATION:  
 APPLICANT: Gil H. Choi  
 TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
 NUMBER OF SEQUENCES: 496  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Human Genome Sciences, Inc.  
 STREET: 9410 Key West Avenue  
 CITY: Rockville  
 STATE: Maryland  
 COUNTRY: USA  
 ZIP: 20850  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.50 inch, 1.44M storage  
 COMPUTER: HP Vectra 486/33  
 OPERATING SYSTEM: MSDOS version 6.2  
 SOFTWARE: ASCII Text  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/071,035  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: A. Anders Brookes  
 REGISTRATION NUMBER: 36,373  
 REFERENCE/DOCKET NUMBER: PB369P2  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (301) 309-8504  
 TELEFAX: (301) 309-8512  
 INFORMATION FOR SEQ ID NO: 358:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1074 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-071-035-358

Query Match 2.0% Score 123; DB 4; Length 1074;  
 Best Local Similarity 18.3% Pred. No. 0.01; Mismatches 415; Indels 352; Gaps 52;  
 Matches 206; Conservative 150;

UY 240 RYRNTAVARAAHLAFDNEHGAVALPPDIITYYFQSSSSGTTTARARRNDVNSTSKPS 309  
 67 RTSLIAYEYNGAKVTFVC-IEPGVSIPTVEVTHGY-----QKNPL 104

UY 410 PSQFERELASIM-AAOTLHAEEVFNNGIYEE-----TPTDIKEMPMFI 353  
 105 PSMISAKALVSVLMKAGDIDITNMVAOKMWEVNGYKLIHSIKRLGASVYIK----- 158

UY 454 GMEGLLPRLNALGSTAR-----VAVYIG-AMVSPNSALYTFEVEDSGMTEKDGAP 405  
 159 SLECKINK-AIEFYKKRPSFHTNTVKTITIGOSTTLLIKNELNISEPKVYVNTA----- 211

UY 406 GPSNRFPYQACPHLAANPQTDGCHVLSSTGSSNTEFSVDYALALGCFAPLARI 465  
 212 NIDRVIGNOLVLP-----NSNSKSGTLLTKKSAGTCTP-VAAKK 251

UY 466 FYLERDGAFTGNGDAIKVYTFPTESEIPGSLCEKHTPRVCAHTVHRLQRMPFPG 525  
 252 AGLOTVMAGALDKPTVAKINVELTKGSLKIKIDKESGDIPEVTEVL-----DECK 304

UY 526 A-TRQPIGVHGINNSOVSQDPLGNA-----PYLLIKRIGDQTEFAKAKM 570  
 405 ALPSKDVTTDKKTL-----STIDGIPHTKVITTEKSVPPYMTIDTPMAATIKAGETI 358

UY 571 QUTYATLEKLFIDLEJURLDKRAGPSSSELS-----SVVDHPT--PRILLIDIRA 621  
 459 SMISKNNHOKODILEKKG-VHTGIDIMNDYSLAGNIFAIKROSPACHVOEITIDKRG 417

UY 622 RLEDTTFPMVLETRQYKINEGSEATHSMALTFDP YSAFCEPTINELVKT 675  
 418 RAE TPKEILANALEIYGYVTE TKSSNFEVNTFPTKVELKVAQVAVLTGTVNKGQ 473

UY 676 HIAVVDIAL-----SUGCHVFYGOOV-----FGRNRKRPQVPLARRFVPL 717  
 474 NOETIGETTLREKDTUNESOGKAEPGAEVTLTARDDGAVKSEAFK-----TEL 526

UY 718 FNRCEITSTRITVILSH-GVSAFNPILAQ-----DAFAGTFEGDILAKVSE----- 764  
 527 VKIILKASLIIVLIALGNNVAVNNALINLFWOLNAPHOLILDBLITVYSIKKVVNN 586

UY 765 ---VIRDIRKKNRV-----FSGNCTNLSENA-----RRLVGLASAVYQREKR 805  
 587 KNAVITRDVTAKEQVIRPFPFEKFNAGSDTAETGFNDLSEKVSPIREITETGAEDKA 646

UY 806 VDMJHGAJGF-----LLKOFHGLPFRGM-----PNSKSK-----PNP 838  
 647 TTANNGOULGFMKGFENLPYGDYILEIEEA---PEGFKITPPEIKSTFEKNKDYAKS 703

UY 819 GMPWTLLQORNU---MPADKLTHEBITTAAVKRFEVEYAINFINLP-PTGIGE 888  
 704 EYVETIIEEDOKPIKNAVIVYEKILNNE---FVSAINLMLYDIDPEKEDSLTS 754

UY 889 LAQFYMANLILKYECHSOYLINTLSITITGARPRDPSSVLIMIKD---VTSADIETQ 945  
 755 LATMKDKNKKINTLDETE LVYDKL-----RYNLHEIKEDWYVAQAADIVEA- 799

UY 946 AKALIEKTENPREMTIAFST-----HLYRAAMNORHVVAGISISKYACA 992  
 800 TKAGDEREKAKPVIAETATIANKEKTKGIMKLIHKLTAEG-----VLKSIYLFVNV 853

UY 993 ANNNRVQANMNSILNPKKNVPLITTEPTPRTIACFPGGCFIPVYTOPSSGNRETTISD 1052  
 854 YENKVAFAEAGNR-----PVA-----KIASLIN 875

UY 1054 GVRGILVSGAMVOLAIYATVAVAGAR-----AGHMAFDWLSTLDEFLARDL- 1102  
 876 QAV TYVCTTERHVSIOFKALHEDGSOFTHGVDWMDVDVSVTHIDVLD 923

UY 1103 FRLHIOIUTLEETWIVGALFAVKIIDE---KITAGD---GEPPTINAFNPD 1149

DB 924 GSKRAEVLITLALLHUCINKLIMKSKIRHHYVNDKPTINIVAAKVDCKYTVCKTFP 982  
 OY 1150 SCEPSIDTTSNV-----LNISGNSISGTVPGIKRPEDE 1185  
 DB 983 -TEIYNEKDNVNGKNHEDLKRSQTLTPKEVPTIPSTIKQPE 1024

## RESULT 6

US-09-071-035-394  
 ? Sequence 394, Application US/09071035  
 ? Patent No. 6448043  
 ? GENERAL INFORMATION:  
 ? APPLICANT: GIL H. Choi  
 ? TITLE OF INVENTION: Enteropneustes laevis Polymorpholides and Polypeptides  
 ? NUMBER OF SEQUENCES: 496  
 ? CORRESPONDENCE ADDRESS:  
 ? ADDRESSSEE: Human Genome Sciences, Inc.  
 ? STREET: 9410 Key West Avenue  
 ? CITY: Rockville  
 ? STATE: Maryland  
 ? COUNTRY: USA  
 ? ZIP: 20850

COMPUTER READABLE FORM:  
 ? MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
 ? COMPUTER: HP Vectra 486/33  
 ? OPERATING SYSTEM: MSDOS version 6.2  
 ? SOFTWARE: ASCII text  
 ? CURRENT APPLICATION DATA:  
 ? APPLICATION NUMBER: US/09/071.035  
 ? FILING DATE:  
 ? CLASSIFICATION:  
 ? PRIOR APPLICATION DATA:  
 ? APPLICATION NUMBER:  
 ? FILING DATE:  
 ? ATTORNEY/AGENT INFORMATION:  
 ? NAME: A. Anders Brooks  
 ? REGISTRATION NUMBER: 36,373  
 ? REFERENCE/DOCKET NUMBER: PB369P2  
 ? TELECOMMUNICATION INFORMATION:  
 ? TELEPHONE: (301) 309-8504  
 ? TELEFAX: (301) 309-8512  
 ? INFORMATION FOR SEQ ID NO: 394:  
 ? SEQUENCE CHARACTERISTICS:  
 ? LENGTH: 1074 amino acids  
 ? TYPE: amino acid  
 ? STRANDEDNESS: single  
 ? TOPOLOGY: linear  
 ? MOLECULE TYPE: protein  
 ? US-09-071-035-394

Query Match 2.0% Score 123; DB 4; Length 1074;  
 Best Local Similarity 18.3% Pred. No. 0.01;  
 Matches 206; Conservative 150; Mismatches 415; Indels 352; Gaps 52;

UY 250 RYRNTAVARAAHLAFDNEHGAVALPPDIITYYFQSSSSGTTTAKAKRNIVNSTSKPS 309  
 67 RTSLIAYEYNGAKVTFVC-IEPGVSIPTVEVTHGY-----QKNPL 104

UY 410 PSQFERELASIM-AAOTLHAEEVFNNGIYEE-----TPTDIKEMPMFI 353  
 105 PSMISAKALVSVLMKAGDIDITNMVAOKMWEVNGYKLIHSIKRLGASVYIK----- 158

UY 454 GMEGLLPRLNALGSTAR-----VAVYIG-AMVSPNSALYTFEVEDSGMTEKDGAP 405  
 159 SLECKINK-AIEFYKKRPSFHTNTVKTITIGOSTTLLIKNELNISEPKVYVNTA----- 211

UY 406 GPSNRFPYQACPHLAANPQTDGCHVLSSTGSSNTEFSVDYALALGCFAPLARI 465  
 212 NIDRVIGNOLVLP-----NSNSKSGTLLTKKSAGTCTP-VAAKK 251

UY 466 FYLERDGAFTGNGDAIKVYTFPTESEIPGSLCEKHTPRVCAHTVHRLQRMPFPG 525

```

Db 252 AGLOTVMAGALDKPNTYAIKINIVETKGS-LKTIKKIDKESGDIYVETVFL-----DFGK 304
OY 526 A-----TRPIGVPCYMSNOYSDICDPLGNVA-----PYLLIRKQGDTEAKAKM 570
Db 305 ALPSKDVITDKDGI-----SLIWKIPHCYKVIITFKSVDPDPMITPTTMAAATKAGETI 358
OY 571 QDTYATLTERLFTDLEOERILNRCAPCSSEGLS-----SVIVDHPT-----FRILDLRA 621
Db 359 SMISKNMKKQGLIILKIC-VEIGITILMINDNLSLACNTFAIRKDSACELVCEITITDEKG 417
OY 622 KILQITTCMKVILVEIKDKIRKGLSEATHSMALTPDP-----YSGACPTTNPLVKRT 675
Db 418 RAE--TPKELANALEGTYYTE--TKSSNGEVNTEKPTKVELKYANQVVALVTSNWKQ 473
OY 676 HLAAYODIAL-----SOCHCYVYQOV-----EGKNFRNQOPLYLRKRPVLD 717
Db 474 NOHITGETITTKEDKDPYCNESOKAFKCAEYVLTAKDQAAVMSBAR-----TEL 526
OY 718 FNGFISTKSTVTLSE-GPVSAFNPILG-----DAPARTFGDGLARVSE-----764
Db 527 VGTASODETVTLALDKKNOVAAYKHLAINFEMQETKAPGCTLDFTKYPVSIKKYDNNE 586
OY 765 ----VIRDIRKVRNV-----FSGNCTINLSEA-----RAKLVGLASAYQOEKR 805
Db 587 KNAVITRDYTAKEQVITFPTFEKFAASALGTAEFTGNDLSFKVSPLEGTXTITGAEDKA 646
OY 806 VMHICALG-----LKOPHGLLPFGM-----PPNSKS-----PNP 838
Db 647 ITACNQLOLFDGCKFENLPYGYLLHEIRA---PECFQKITPLEIRSTFKKKDDYAKS 703
OY 839 QWETLLQBNQ-----MPADKLTHEEITITAAVKRTEEYAAINFILP--PTCIGE 888
Db 704 EVFETTERGQKQPIKMWIVPYEKIINN-----FVSLNRLMLYDLPEKEDSLTS 754
OY 889 LQOFYMANILKQCDOSVYILNLTLSITGARRRPPSSVLMHMRK---VTSADILFTQ 945
Db 755 LKWKQGNKKLNTLDTE-LVDKL-----RYLHEIKEDWYVAAQIDVIA 799
OY 946 AKALIEKTNLPRLMTAFTST-----HLVRAAMNORPVNVLGISKYHGA 992
Db 800 IMAQCKDKKAKIVVIAETAILANKRKITGIMKILKITAEO-----VLKSTVLINNV 853
OY 993 ANNNVPEVQNMKSLNKGKVVPLFTFTTRTRFTIAPPRGFIQPVTPGSSGNRETLSD 1052
Db 854 YENKVAEAGNE-----PVA-----KDAALNN 875
OY 1053 QVGGIIVSGAMVOLAIVATVRAVGAR-----AQHMAFDDMLSLJDEFLARDL-- 1102
Db 876 QAO-----TNCITIERHVSIGOTKAHLEDGSGOTFHGVMMDFDVSATHVDL 923
OY 1103 --ELHDOIIOTLETTWTEGALBAVKTLDE-----KTTACD-----GETPTNLAFNFD 1149
Db 924 GSRAREFTILYALIPDILNKEIMKSKTEHVVNDKFKTIVIAKVKVIGYPRGKIFL- 982
OY 1150 SCEPSHDITSNV-----LNISGNSISGTVPLKRPPEDE 1185
Db 983 -TEINTEKQCNVKNCKHNDIKKESQITTPKKEVPTTIPSTPOPE 1024

```

```

; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 5215
; TYPE: prt
; ORGANISM: Streptomyces venezuelae
US-09-105-537-2

Query Match      1.8% Score 111 Db 4 Length 5215:
Best local Similarity 19.7% Pred. No. 3.4 419 Indels 494 Caps 66:
Matches 250; Conservative 109; Mismatches 411

OY 51 NLTEKFTFTSLAVV-----SGAFTTGLAVAGTITKLTTSHPYSPVVFHSGKHLP 102
Db 166 SLTVDAQSSSLVAHILACESLRAGESTALV-AGVNL-----NILA 206
OY 103 SSAAPNLTACNARARHFGFSGCGPPVDGAVETGCAFTCTRIQLEPENTILVIALP 162
Db 207 ESA-----VTEERFG-----GLSPDGTAYTFDARA-- 231
OY 163 KEAVFMGNVFLHVGGLDIVIHINQDVIRIPFLFVQLEMPVNRLVPDPTIHRSIQEGF 222
Db 232 -----NCFVRGEG-----GSVVVLKPL-----SRAIACDQVH-----CV 261
OY 223 VYPTPEYNTGLCHLHDCVTAIPMAVALPVKNVTAVARGAHLAFDENHCAVLPDITTYT 282
Db 262 IRASAVNNDG-----ATPGLTVPSSRAOGEKVLREAYRRKAALDPS-AVQ 303
OY 283 YPOSSSGCTTARGARRNDVNSTSKPSPSGCFERILASIMADPTALHAFVIFNTGVEET 342
Db 304 YVELHGTGT-----PVGDPIEALAGVALSARADPL-----LVGSA 342
OY 343 FTDIKEMPEFIMGECTPLRLNALGSYARVAVIGAMVFSNSALYTEVEDSGMTEAKD 402
Db 343 KTVNGHLEGAAGIVGLIKTLTALG--RRRIPASLNFRTPHPDIPL-----DTGLIIVPD 394
OY 403 GGRSPSPNRYQFAPGPHLANPQTDKD-----GHVLSQSTSSSTEESSVD 448
Db 395 G-----LREMPHPDRELLAGVSSFGMGKINAHAVLSBGPAGGEGQPGID 438
OY 449 YIALICGPGATLARTLFYLERCDAGA-----FTGCHQDAIK-----YVGTPTDSELP 496
Db 439 EETPV-----DSGAALPFEVYGRGGEALRAQARLHEAVEADPELAP 480
OY 497 GSLCEK--HTRPVCAHTTV-----HRLPQAMPFPQATRPPTVYFGIMNYSQDCTPLG 548
Db 481 AALARSLVTRTFVTHRSVYLAPDRARLLDGLAAGTAPAGVVTG----- 528
OY 549 NYAPYLILRKPGDQTEAAKATMDTYRATLERLFTLEOERILDRGAPCSSEGLSVYD 608
Db 529 -----PAG-----RLAVLF-----SSQAGORTGMG-ELYNA 555
OY 609 HPTFRILDLRLARI EOTTTOFMKVLVETRD-----YKIRGLSEATHSMALF 656
Db 556 HPAFAIADFADAVALDPLDRLPRLAELVAAGDTLDRVTHTTOPALFAVEVALRLVESWCVT 615
OY 657 FDPVSGAFPTTNFLVNRHLAVVQDLALSGCHVFYGLVGEIRFNPQOPVLPRPREVD 716
Db 616 PDLLAGH-----SVEISAHYAGV--LSLRDA-----ARLYAAGRLMALP----- 656
OY 717 LENGFISTRSITVTLSEGPSAPNPITLGODAPGRTFGDGLARVSEVIRDIRKRVN 776
Db 657 --EGGAM--VAVEASEEVL--PHL-----AGREBELSLAAVN-----GPRAVV 694
OY 777 FSGNCTINLSEAKARILVGLASAYOROEKRVDMHICALCFLIKOPHGLP----- 825
Db 695 LAG-----PR-----GMPNKSPPNQW-----FWTLQBNOMFADKLTHEEITIAV 866
OY 826 -----PR-----GMPNKSPPNQW-----FWTLQBNOMFADKLTHEEITIAV 866
Db 742 EELDPQERHVVVSVIAGLAPYTA-----GQWIDPXYW-----VQGV--RRPVAFILAV 787
OY 867 KPFTEEYAINFINIPP--TGTIGLACFYMANLILKYVDHISQYLINTLTSTTTTAPPPRO 924

```

DB 788 KETTESCAPILEEDHGVNSAMADSV-----KIDJAAVAVASLRKG--RP-E 833  
 QY 925 PSSVLMIRKDVTSAMIEUOKALLEKTENLPELMTAFTSTHLVRAANQRPVVLGI 984  
 DB 844 POSTLAAITTVYKRIQVD-----WTAAKSGTAVPL---PTVAFOR 874  
 QY 985 SISKYGAACNNHVFOACNNKCKNGCVLFFPDTRKPLIACVRCGICVYCHSSG 1044  
 DB 875 ERIMFKAATAAFLTAGR SGTAG-----TGFAG 905  
 QY 1045-----NKETLSQVRGIIVSRCAMQVLAIVATVVA-VARAA 1081  
 DB 906 VNSGEGCHETKAGACGQDPAHETT--KRAHAAVA-----VLKDDPTVVEIGLIF 957  
 QY 1082 QJMAFPDMIS-----LTDEFLAKDELELHQI10TLETPTVGEALVAVKILDEKTTA 1135  
 DB 908 KELGDSLSMVEIKNALVDTGLRLPSGLLTFH-----PTP---PALAA--HLLGELTIG 1006  
 QY 1106 GKEETTNALPNDSCPS--HDTSNVANI SCSNISCSTVPGIKRPEDDELFDL----- 1189  
 DB 1007 GSGCTG--SMLGTPATPATDIAEPALAIIG---MACRYPGSVISPED--LMRLVAG 1056  
 QY 1190 SCPIKING 1197  
 DB 1057 KDAVSGLPITKRG 1068  
 RESULT H  
 DB 07 642 744C 4  
 Sequence 4, Application 05/07642744C  
 Patent No. 5824513  
 GENERAL INFORMATION:  
 APPLICANT: Katz, L.  
 APPLICANT: Demadio, S.  
 APPLICANT: McAlpine, J. H.  
 TITLE OF INVENTION: Recombinant DNA Method for Producing  
 TITLE OF INVENTION: Erythromycin Analogs  
 NUMBER OF SEQUENCES: 27  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Edward H. Gorman  
 STREET: Abbott Laboratories D377/ABD-2 One Abbott  
 CITY: Abbott Park  
 STATE: IL  
 COUNTRY: US  
 ZIP: 60064-4500  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC DOS/MS DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: 05/07642744C  
 FILING DATE: 17 JAN 91  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Danckerts, Andreas M.  
 REGISTRATION NUMBER: 43652  
 REFERENCE/DRAWING NUMBER: 4952.US.01  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 708-947-9396  
 TELEFAX: 708-948-2624  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3567 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 DB 07 642 744C 4  
 Query Match 1.78; Score 106.5; DB 2: Length 3567;  
 Best Local Similarity 19.98; Pred. No. 4.8;  
 Matches 237; Conservative 104; Mismatches 384; Indels 465; Gaps 62;

QY 125 CQGPVVDGAVETTCAR--ICGR-----ICGEPENTL-VIVTALPKRAVENCNPLH 174  
 DB 1165 CPGDDLLAAVEBAGASVAVVQAQDAALREALGDEP-----VTAL-----VII 1205  
 QY 175 YGGDIDVHINHGIVTKIPLPPVQLFMDVYKMLVNDPNTTHRSIGRCVYVTPPYNTIGC 214  
 DB 1206 AGTL-----TNGC-----SISVADERP-----ATIIAAKITAL 1214  
 QY 235 HLLHDCVIAPMAVALRY-----RNVAVAPQAAHL-AFDEHECAVLPDITTT 282  
 DB 1235 AVI--DEVLGHPAARPRVYVSSVACIWDGACMAAAYAAQSAVIALAPRHR----- 1282  
 QY 283 YFOSSSGCTTACARANDYNSTSKPSPG-----GF--EKRLASIMADYALHA-EVIFN 335  
 DB 1283-----ARGRSTSAVMTPMALPGAVDDGTLREGLRS-LSADRAMKMEVLA 1330  
 QY 336 TGIYEELPDIKEMPMH-I-GMEGLTPRLNAGSYTANVAGVIGAMVSPNSALYIIEVD 344  
 DB 1331 AGHVSAVADV-DWPLYISGCPAATRP-----TALPAPIAGR 1365  
 QY 395 SGMTEAKDGSGPSFNFYFAG-----PHLAANQTDREDGIVLSSQSTGSSNTE 444  
 DB 1366 GQQAFAAPDS-GPTGAPRAQRLAGLSPDQDNLLEIVANVAAE--VIGHESSAEINVR 1420  
 QY 445 PSVDYIALICGFGAPILARLLFYLERQDAGAPIGCHQDALKYVGTPTDSIPGSLCEKHT 504  
 DB 1421 RARSELGLDLSLNMALRKRL-----SASTG-----LPLPSL 1452  
 QY 505 RPYCAHTTVKTKG--HMPH+GALIKYV+GYS+GMS+YSLCPL-----GNYA 551  
 DB 1453--VFHDPTVYALAOHLKARLVGDAIDAAVRYGAAD--ESHPILAIVGICGRPGIGS 1506  
 QY 552 PYDLR-----KPGDQTEAA----- 566  
 DB 1507 PROLMKVIAGANITTCFPARQDQDQMLHYPRDNGTGYVKGCTITLAAIDPDTPFG 1566  
 QY 567-----KATMODYRATLERLFDLEGERLIDKCAPSSKGLSSVIVDPTFR 614  
 DB 1567 TTPREAIAMDPOORLMELETAMFAVERAFTDPAIFGTTVFVFMNGSQVQLAVEAE 1626  
 QY 615 I-----LDTLIAH+PQ-----TQPMKVIIVETNDYKIRGLSIAHSMAI 655  
 DB 1627 VDGVOGIGNSASVLSCHIAVTFGMKQALIVDTACSSSLVGI--HIAOMALRKQDLSAL 1684  
 QY 656 I-----FDYSGAFCPITNFIYKRTHLAVVODLALSOCHVEYGOVGEGRNFEN 704  
 DB 1685 AGCVTMSDPYT-----FVDFSTOR-----GIA-SIDCKKAFSAADGFALSHQVAAI 1731  
 QY 705 QHPYVILKKHVVILHNGHISTKSTIVILSHQVSAVNTICQAPAG-----RTFGHILAR 760  
 DB 1732 VLEPLSRAR---ANG-----HGVLAVLRGSAAVNGPNSGLAAPNSGSEVIRQALAA 1782  
 QY 761 VSEVEILDIKRVKNRVVSCNCTNLSAARARIVGLASAYOROKR-----VDMH--- 810  
 DB 1783 SGVPA-ADVVV--VEAHGIGTELQDPIHAG--ALLATVGDMDRLIKASVNTINIGHTO 1836  
 QY 811-----GALGELLKUFHGLLFPKGMPPNSKSPQWMTFLLQBNQMPADKLTHEETTIA 864  
 DB 1837 AAGAGAGVILKVVILAMRHGM-L-PSLIHADELSPHIDW-----ESGAVVILKEEVWPA 1887  
 QY 865 AVKRFTFEYAAINPILNPICIGFIQOFTYMANILKYCHISOVLINWISITIGCARPRD 924  
 DB 1888 GER---PPRAGVSSFGVSST-----NAIVIVEAPAEQ-----EAARTEKG 1925  
 QY 925 PSSVLMIRKDVTSADIEUOKALLEKTENLPEL-----WTATTSIHIL-VKRAMNOR 977  
 DB 1926 PLPFLVLSGSRSEAVVAA--QARALAHHLADTPHIGITTAAMTALACRAAFVYKAAV---- 1978  
 QY 978 PMVVLGISTISKTHGAACNNHVFOACNNKCKNGCVLFFTPPTPPIIAPPGSFLICP 1037  
 DB 1979-----IGDPRASVCAFLDALAPR-----PSAAVAVAP 2005



QY 1036 VTGSSGNKRETTLSVGRILVSSANMVLATVAVAGARAHMAFDMJLTLDEEF 1097  
 DB 2006 VT---SAPRKPYL-----VPPGCAO-----WVC-----2026  
 QY 1098 LARDLELHDIQTL-----ETWVEGALFAVILDEKTAGAG 1138  
 DB 2027 MARDLESSEVFAESMRCAELSPHTDM-----KLDD--VVRGDS 2065

RESULT 9  
 US-08-439-009A-4  
 Sequence 4, Application US/08439009A  
 Patent No. 6004787

GENERAL INFORMATION:  
 APPLICANT: Donadio, S  
 APPLICANT: Katz, L  
 APPLICANT: McAlpine, J H  
 TITLE OF INVENTION: Method of Directing Biosynthesis of  
 TITLE OF INVENTION: Specific Polypeptides  
 NUMBER OF SEQUENCES: 27  
 CORRESPONDENCE ADDRESS:  
 ADDRESSER: Steven F. Weinstein  
 STREET: Abbott Laboratories D377/AL6D-2 One Abbott  
 CITY: Abbott Park  
 STATE: IL  
 COUNTRY: US  
 ZIP: 60064-3500

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-MS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/439,009A  
 FILING DATE: 11-MAY-1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Casuto, Dianne  
 REGISTRATION NUMBER: 40,943  
 REFERENCE/DOCKET NUMBER: 4952.US.D1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 847-938-3137  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3567 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-439-009A-4

Query Match 1.7%; Score 106.5; DB 3; Length 3567;  
 Best local Similarity 19.9%; (Prod. No. 4.8;  
 Matches 237, Conservative 104; Mismatches 384; Indels 465; Gaps 62;

QY 125 CGSPVAGAVETGAE--ICTR-----LGEPEPTILLYLTALFKRAVFNKVFELH 174  
 DB 1165 CPDIDILAAVEFAGASAVVCAQDAALREALGDEP-----VTAL-----VH 1205  
 QY 175 YGGLDLYIHNGGVIRLRPLRVQULFMPDVNRLVDPREYFNTHKISGEGFYVTPREYNTGLC 234  
 DB 1206 AGTL---TNFG-----SISEVADEEF-----AETIAAKTALL 1234  
 QY 235 HLHDCVLAIPMAVALRV-----RNVAVARGAHL-AEDENHEGAVLPDITTYT 282  
 DB 1235 AVL-DEVLGRAREVREYSSVAGICGACMAVYVAGSAYLDALAEHHR----- 1282  
 QY 283 YFOSSSSSGTITAKAKNDVNSTSKRPSG---GF---ERLASIMADVALHA-EVIFN 335  
 DB 1283 -----AKGRSCTSYAAMPWALPGCAVDGDLREKGLRS-LSADKRMKRTVERVLA 1330  
 QY 336 TGIYEETPTDIKEPMFI-TMEGTLPLKLNALGSTARVAGVIGAMVSPNSALYLTVEED 394

DB 1331 AGPVSAVAVADV-DWPEYLSGSEFATRP-----TALFAELAG 1365  
 QY 395 SCMTFAKDGPGPSEFNRYOFAG-----PHLAANPDTRDGHVLSSTGSSNTE 444  
 DB 1366 GGAFAEAPDS-GTGPAPRALAGLSPDEOENLLLELVANAVAR VIOHSAARINVR 1420  
 QY 445 FSVYDALTCGFAPLARLLFLYLERDAGAFGTGSHDALKYVTGTFDESEIPLSLFEKHT 504  
 DB 1421 RASELGDSLNAMALRKRL-----SASIG-----LKLPAKL 1452  
 QY 505 RPYCAHTVYHRLRO-RMPFPGATROPICVPGTMSQVSDCDPL-----GNYA 551  
 DB 1453 --VEDPVTYALAQHLRAPLVDADQAAVYVGAAD---ESEPIAVISIGTFEPGATIS 1506  
 QY 552 PYLLR-----KPGQITRAA----- 566  
 DB 1507 PROLMVLAEGANLTTGFPADKWDIGRLYHPPDNDPGISYVKGCHLTIDADPDGPHG 1566  
 QY 567 -----KATMDYRATLERLEIDLEQEPILDPGAPCSSELSVIVDHPFR 614  
 DB 1567 ITRFALAMPDQORLMLETFMWEAVERAGIDPDALKGTDTGVFGMNGQSYMOILACRAFR 1626  
 QY 615 I-----LDTLRARIPOT-----TTQPMKVLVETRDYKIRPLSATRSMAL 655  
 DB 1627 VDGYGUNASVLSGRIAYTFGMEGPAITVDIACSSSLVGI--HLMQALRRGCSLAL 1684  
 QY 656 T-----FDYSGAPCPITPFLVKTHLAVVUDLALSUCHCVYGGQVQVPRKRNFN----- 704  
 DB 1685 AGCVTWASDPYT-----FVDFSTOR-----GLA-SDGCKAPSAADDPALSKCYAAL 1731  
 QY 705 QFQPVLRFRFVDELFGSFISTRITVLSGFSAPNPILGQDAFAG---RFFDGLAR 760  
 DB 1732 VLRLPLRAR---ANG-----HQLAVLRGSAVVDASNCIALAPNQPQSRVIRQALAA 1782  
 QY 761 VSEVIRDINVKNRVYSGCNTINSEARARIVGLASAYQROEKR-----VDMHL-- 810  
 DB 1783 SGVPA-ADVYV---VEAHGTGLGDPTEAG--ALITTYGQDRDRPLRLGSKTNIGHQ 1836  
 QY 811 -----GALGFLKQFGLFLPRGMPNPSKSPNPOWFTLLDRNQMADKLTREITTTIA 864  
 DB 1837 AAGACAGVIVKVLAMRHGML-PRSLHADDELSPHIDW-----ESCAVEVLRREVWPA 1887  
 QY 865 AVKRFTEYALINFINPPTICIGLAQFYMANLILKCDHSQYLINLTLSITARRPRD 924  
 DB 1888 GER---PRRAGVSSFGVSGT-----NAHVIVEAPAHQ-----FAAKTEKG 1925  
 QY 925 PSSVLHWIRKDVTSADILEQAKALIEKTENLPCL-----WTTAFISTHL-VRAAMNOR 977  
 DB 1926 PLFVLSGRSEAVVAA---QARLALHLRTPDELGLTDAMWTLATRRAPFDVRAAV--- 1978  
 QY 978 PMVVLGISISKYHGAAGNNRYFOAGNNSGLNGKNNVCHLFTTRTRRPIIACPRGCFICP 1037  
 DB 1979 -----LDUDFAGVCABLLALAEFG-----PSADAVAP 2005  
 QY 1038 VTGSSGNKRETTLSDYRGILVSSGAMVULATVAVAGARAHMAFDMJLTLDEEF 1097  
 DB 2006 VT---SAPRKPYL-----VPPGCAO-----WVC-----2026  
 QY 1098 LARDLELHDIQTL-----ETWVEGALFAVILDEKTAGAG 1138  
 DB 2027 MARDLESSEVFAESMRCAELSPHTDM-----KLDD--VVRGDS 2065

RESULT 10  
 US-07-731-157A-7  
 Sequence 7, Application US/07731157A  
 Patent No. 5457032  
 GENERAL INFORMATION:  
 APPLICANT: Quak, Wilhelmus J.  
 APPLICANT: Misec, Onno  
 APPLICANT: Van der Laan, Jan M.  
 APPLICANT: Lentling, Herman B.M.  
 TITLE OF INVENTION: Mutated beta-lactam acylase genes

```

NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSER: COOLEY GODWARD CASTRO HODDLESON & TATUM
STREET: FIVE PALO ALTO SQUARE, 4TH FLOOR
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/741,157A
FILING DATE: 19910509
CLASSIFICATION: 415
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 90200962
FILING DATE: 18 APR 1990
ATTORNEY/AGENT INFORMATION:
NAME: RAE-VENTER PH.D., BARBARA
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: GBK0-027/0005
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415 494 7622
TELEFAX: 415-857 0663
TELEX: 480816 COOLEY PA
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 774 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Pseudomonas species
STRAIN: SE83
US 07 741 157A-7

Query Match
1.7% Score 106; DB 1; Length 774;
Best Local Similarity: 28.2%; Prod. No. 0.31;
Matches: 145; Conserved: 79; Mismatches: 251; Indels: 242; Gaps: 37;

439 AACAAATADENHAKAV LPPDIIYTYOSSSSGTTAGAKRNHYNSTSKISVSGGPER 316
184 AANALIKLRYDKGQDILTPGVFAERLEADIALRPVADALLKMGCGASJAGCGSNN 242
417 RLAS---LMAVDLALAEVFTGTYEETPTDIKEMPMFTGMEGTLPRVALGS 367
241 MAVAVRTATAGPITLACDHPPEPT---PGMYAQHHLACDHPDM-IGL---IV----- 289
468 VYARVACVIGAMVFSSNSA-----LVLTEVHDSQMTF----- 399
290 ---GVGPFPEFAHNEVVAVVCTIAFMQIHLLYLEGFAEDRRTARFGNFFEPVAMPRO 343
406 AKIKGPGHSNRRYUFGPHLAAMVQDRGCHVLSSTGSSSTNTPESVDYIALICGCG 457
444 KLAIVGAGALREDFVITRRGVATADP---LPCAALILRSVQAFEDLSFDCTIRMP--G 398
438 ATLAKLLEFLERLAGATG GH GLAKVYTGTFDESLPCLSEKHKIKPVLAHT 511
399 ASTVADLY-----DATKMGCLIDHNLVAGDVAGSIGHVYRAVAVSRPENCMLVYPCMS 452
512 IVHRLR---GRRHFGCATKPPICVCTMNSQY-----SDCHDLCVAVY--- 553
431 GEHEKGMVPIHEAMP---VIDPPEGLIVTANNRVVADDPYALCTDCHP---PYRAE 504
564 ---LILKKPGIOTFAAKATMOPT---YRAITERLPI---DLRQERH----- 590
505 KIMERIVASPAVAVIDAAAHATITISPHVGLKARLEAGCIGCSLPAHELKOTLLAMDKR 564
591 LDRGAPSSSESSLSVIVIPTEFRILDTLRKILFQTTQFMRYLVETRDYKIREGISEEA- 649

```

```

DB 565 MDAGSQASAS-----YAFRRAL-----TRLTATSGLEQAI 596
QY 650 TSMALL---FIPYSACCHLITNFKKTHAAVU-----UAIASCHGVYRGQVGRNP 702
DB 597 AHPFAAVPPGVSPGGGVMAVPT-LLRNDAGMLKGSWMDEALSEALSV-ATONLTSGW 654
QY 703 RNOFOPVLRREVDLFGNGCFSTRITVTLSSEPVASP-----NFTLRQDAP 749
DB 655 GRRHRP---RRTHPILSAQPPAMAALL-----NPSKPTGGDGYVLANGVYSAGHPAT 705
QY 750 AGTFEGDLARVSVEVIRKRVNPFVS--GN-TNLSAARPAVLVLSAVYGPQEKVDM 808
DB 706 YG-----ALSRVYFQVGNMDN-----SKWV 726
QY 809 LHCALGPIILKQTHGLLTPKCMTPNSKSPINQW-----FWTLQRNOMPAUKLTHEEL 860
DB 727 FRCASG-----HPASPHYADQNAFWSDCAMVPMLYSMDKILAEAVTSOEL 771

RESULT 11
US-08-541-780-7
Sequence 7, Application US/08541780
Patent No. 5935831
GENERAL INFORMATION:
APPLICANT: Quax, Wilhelmus J.
APPLICANT: Missel, Onno
APPLICANT: Van der Laan, Jan M.
APPLICANT: Lenting, Herman B.M.
TITLE OF INVENTION: Mutated beta-lactam acylase genes
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSER: COOLEY GODWARD CASTRO HODDLESON & TATUM
STREET: FIVE PALO ALTO SQUARE, 4TH FLOOR
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/541,780
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/731,157
FILING DATE:
APPLICATION NUMBER: EP 90200962
ATTORNEY/AGENT INFORMATION:
NAME: RAE-VENTER PH.D., BARBARA
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: GBK0-027/0005
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-494-7622
TELEFAX: 415-857-0663
TELEX: 380816 COOLEY PA
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 774 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Pseudomonas species
STRAIN: SE83
US-08-541-780-7

Query Match
1.7% Score 106; DB 2; Length 774;

```

Best Local Similarity 20.7%; Pred. No. 0.31;  
Matches 145; Conservative 79; Mismatches 251; Indels 242; Gaps 37;

```

QY 258 AKGAHAEHDEHNGAV-LIPVITYIVYSSSGTITAKGAKRNVNISKISPSGGRK 316
DB 183 AANMLKEPYDKNGJULCLIPHSVEAEKLEADLAALPRAVDALLKMGSGASDAAGSNN 242
QY 317 -----RLAS---LMAADTALHAEIFNTGIYEETPTDIKEMPMIGEGTLPRLNALS 367
DB 243 WAAVAGKATIGPILAGIDHKKVYF-LPQYAOHHLAITHTM-LCL-TVP----- 289
QY 368 YTAHVAGVIGAVESPNSA-----LYLVEEDSGMT----- 399
DB 290 -----GVAFEPHFAHNSKVAVCTHAFMDIHLYLEQFAEDGRTARFNEPEVAMPFD 343
QY 400 ---AKDGGCPSPNRYOYAGPHLAANPOTDRGCHVLSSTGSSSTSTSTSVYDIALICGCG 457
DB 344 RIVAGVAMPEDTIVETHPHPIAGTP---LEGAALTLSKSVGAFETDLSFQCLTHMP--G 398
QY 458 APLARLLEYLERCDAFGTG--GH---GDALKYVTGTDESEIPCSCEKHTRPVCAHT 511
DB 399 ASTVAGLY-----DATKGWGLIHHNIYAGDVAGSIGHLYARAVSRRENQMLVYPCMS 452
QY 512 TVHRLR-----QKMPFGQATRPQIGVFGTMSY-----SDCLPLGNVAPY--- 553
DB 453 GEHEWRGTIPHEAMP---VIRPPGGLITVANNRVADDPDYLCDCHP-----PYRAE 504
QY 554 ---LILKKPGQIPEAKATIMQD)-----YKATLERFI-LDEQEPH----- 590
DB 505 RIMELVAVSPAFAVDAAHIAHDTLSPIHGLLRARLEAGIGSLPAEELRQTLTAMGDR 564
QY 591 LDPAPCSSEGLSVYVHPTRFRIIDTLRARFQDTTQFMKVLVETPDYKIPREGLSEA- 649
DB 565 MDAGSOAASA-----YNAFRKAL-----TRIVTARSGLEQI 596
QY 650 THSMALT---EDPYGARGCPTNFIYKTHLAIVO---DLASCHICFYGOQVEGRNF 702
DB 597 AHPEAAVPPGVSPQGVWMAVPL-LLRNDAGMLKMSWDEALSEALSV-ATQNTLTGKAW 654
QY 703 KNOFQPLRRRFRVDLFPNGCFTSTRSTIVTSLGSPVAP-----NPTLGGDAP 749
DB 655 GEEHRP---RTIHPLSNQFPAMALL-----NPYSRPIGSGDGYLANGLPSAGREPAT 705
QY 750 AGRTDGLARVSEVETRIYKKNVYFS-GNCTNLSEARARLVGLANAYQDQEKRVDM 808
DB 706 YG-----ALSRYVFDVGNWMDN-----SRWV 726
QY 809 LHGALGFLIKOFGHLLFPRGMPNSKSPNOM-----FWTLQBNOMPADKLTJHEE1 860
DB 727 FHGASG-----HPASPHYADONAPMSDCAMVMLTSMWRIAAEAVTSGEL 771

```

RESULT 12  
US-08-633-760-46  
Sequence 46, Application 15/08633760  
Patent No. 5804429  
GENERAL INFORMATION:  
APPLICANT: NIMA, MINEO  
APPLICANT: SAITO, YOSHIMASA  
APPLICANT: FUJIMURA, TAKAO  
APPLICANT: ISHII, YOSHINORI  
APPLICANT: NOSHIGUCHI, YUTJI  
TITLE OF INVENTION: A NEW CRYSTALLINE POLYMER  
NUMBER OF SEQUENCES: 64  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ORION, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE 400  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/633,760  
FILING DATE: 01-MAY-1996  
CLASSIFICATION: A35  
ATTORNEY/AGENT INFORMATION:  
NAME: ORION, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 18-929-0 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-5000  
TELEFAX: (703) 413-2220  
INFORMATION FOR SEQ ID NO. 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 774 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-633-760-46

Query Match 1.7%; Score 105; DB 1; Length 774;  
Best Local Similarity 20.7%; Pred. No. 0.39;  
Matches 146; Conservative 76; Mismatches 241; Indels 242; Gaps 37;

```

QY 109 LTR--ACNARERFGSSRCQGPVGVAVETGAE-ICTR-----LGLEPENTILYLVTAI 161
DB 70 LTRKRALGRAAMWLG--AEAAEADILVRLGEMKXCRRDFEALGVLEAD-----M 117
QY 162 FREAVFMCVNFIHYGGLDVIHINHGDIKIRIPVQVLFMPDVRILVDPDPNTH-----K 216
DB 118 LRAVYAGVNAFLASGA-----PL-PVETGLIGAR--PELPEHMHSLAVAK 159
QY 217 STGEGVVTPPYNTGLCHLHDCYIAPMAVALRPNNTAV---ARGAHLAEFENHES 272
DB 160 RLG-----LIMGAVWFKIMRMIALPVGAAMALKLKYDDCGRD 197
QY 273 AV-LPPDITVYVFOSSSGTITAPGARPNVNSTKSPSGCFEP-----PLAS---IM 322
DB 198 LCIPFGAEDRLREADLALTRPVDALLKMGSDASDAAGSSNNMAVAPGRATATGPIL 257
QY 323 AADTALHAEIFNTGIYEETPTDIKEMPMIGEGTLPRLNALSATYARVAGVIGAVS 382
DB 258 AGDPHVFEL---PKYAOHHLAQDRFDM-IGL-TVP-----GVGFPHPFA 298
QY 383 PNSA-----LYLVEEDSGMT-----AKDGGGSPSNRF 412
DB 299 HNGKAVYCVTHAFMDIHLYLEQFAEGRTARPGNDEPEVAMSRDRIAVGADREPDIV 358
QY 413 YQFAGPHLAANPOTDRGCHVLSSTGSSSTSTSTSVYDIALICGFGAPLARLIFYLRQD 472
DB 359 ETRHSGPIVAGDP---ELGAALTLSKSVGAFETDLSFQCLTHMP--GNSTVAGLY-----D 407
QY 473 AGAFTG--GH---GALKYVTGTDESEIPCSCEKHTRPVCAHTTVHRLR-----QMP 521
DB 408 ATIRGWLIIHHNIYAGDVAGSIGHLYARAVSRRENQMLPVPMQSGEHHKRWIPIHAMP 467
QY 522 REGQATROPPIGVFGTMSY-----SDCLPLGNVAPY-----LILKKPDQTE 544
DB 408 K---VIDPFGIIVANNKVVADDPDYLCDCHP-----PYAKERIMKRIVANPAPAVY 519
QY 565 AAKATMONTVPATLFRIFIDLRQERLLDRCA--PSSSEGLSVIV-----D 608
DB 520 DAAAHIAHDTLSP---HVGILRRPLEALGARBDSSAEGIPQMLVAVDGMADAASEVASA 574
QY 609 HEPERRIIDLTL---RARIQTTG-----FAKVIYETRYKIRKQ--- 645
DB 575 YNAFRKALIKLYTRKGLRQALISHPFAAVALGVSPQGVWMAVPLILKDDDACMLKQMSW 634
QY 646 ---LSEA-----THSMALTDPYSGAFCP 666

```

DB 635 DQALSEALSVASQNLIGRSGWGEHHRPRFTHPLATOPAMAGILNP 679

RESULT 14

US-08-633-760-48  
Sequence 48, Application US/08633760  
Patent No. 5804429

GENERAL INFORMATION:

APPLICANT: NIMA, MINO

APPLICANT: SATO, YOSHIMASA

APPLICANT: FUJIMURA, TAKAO

APPLICANT: ISHII, YOSHIMORI

TITLE OF INVENTION: A NEW CEPHALOSPORIN C ACYLASE

NUMBER OF SEQUENCES: 64

CORRESPONDENCE ADDRESS:

ADDRESSEE: OHLON, SEIVAK, McCLELLAND, MAIER & NEUSTADT,

STREET: 1795 JEFFERSON DAVIS HIGHWAY, SUITE 400

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/633,760

FILING DATE: 01 MAY 1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: OHLON, NORMAN F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 18-929-0 PCT

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

INFORMATION FOR SEQ ID NO: 48:

SEQUENCE CHARACTERISTICS:

LENGTH: 774 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-633-760-48

Query Match

Best Local Similarity 1.7%, Score 105, DB 1, Length 774;

Matches 146; Conservative 76; Mismatches 241; Indels 242; Gaps 37;

DB 109 LTR ACNANAREGEFSGGPPVGNVFTTGAETCTP---LGLPENTILLYVTAL 161

DB 70 LTRKALDKAEMILG AFAAENDILVRLGMEKVGCRDFEALGVKND-----M 117

DB 162 FKEAVFMGNVLTGCGDILVHINHGDIKILFVQLFMPVNNKLVDPPTNTH-----R 216

DB 118 LKAVVAGVNAELASCA-----PL-PVEYGLGAE--PEPWEMISIAVMR 159

DB 217 SFGGEVYVTFEYNTGTHLHLPVIAPMVALVRNTAV---ARKAHLAFENIEG 272

DB 160 KLG-----LMGSAVPMKIMRMALPVCAANALKIYIDGGRD 197

DB 273 AV LPPDTTYFSSSGTTTARGAPPNVNSTSKPSPSMEFER-----RLAS---IM 322

DB 198 LKCTPGVAMKRLADLALTRPVAILLKAGSGDAAGGNNNAVAPGTATGRTIL 257

DB 424 AADPAIAAEVIFNIGVETITPDIKEMPTOMGCTIPRLNALGSIYAVACVIGAMVS 382

DB 298 AGDPRVFEI---EGIVAQIHLACIDREFD-IGL--TVP-----GVPEFPHFA 298

QY 383 PNSA-----LYLTVEEDSGMTE-----AKDGPGSPENRF 412

DB 299 HNGKVAVCYTHAFMDIHLYLEDFAGEGRARPCNDPEHVAWSRDIYAVGADREFDIV 358

QY 413 YQAPGPHLAANOQDNDGVLVSSQSSNTSPSVYIALICGFGAPILARILFYLERGD 472

DB 359 KTRHGVIAGDP---RDGAALTLRSVQFAETDLSFGLTRMP---GASTVAOLY-----D 407

QY 473 ACAFTG--GH-----GALKVYVGTPIFSEIPCSICEKHTRPVCAHTIVHRIK-----QRM 521

DB 408 AIRGWLIDHNIIVAGDVAGSICHLVVARVPSRPRENGMLVPGMSCHRWKGIIPHAM 467

QY 522 KFSQATKQPIGVGTNNSQY-----SPGDDPGNAVRY-----LIRKPPDQGE 564

DB 468 R---VIDPQGILIVTANNVAVADHDIDYICTDHP-----PYAKRKIKLVANPAVAD 519

QY 565 AAKATMODYRATLERLFIIDLEORILLDQCA--PCSSPGISSVIV-----D 608

DB 520 DAAIHADILSP-----HVGLLRRLKLEALGARDDSAAEGLRQKLVAMDGMDASEVASA 574

QY 609 HPTFRILDTL---PARIROTTO-----FMKVIYTRDKIRK--- 645

DB 575 YNAFRALTRIVTONGSLHQAISHPPAAYAPGVSPGQVMAVPTLLKDDAGMKIOWSM 634

QY 646 ---LSEA-----THSMALTDPYSAGFCP 666

DB 635 DQALSEALSVASQNLIGRSGWGEHHRPRFTHPLATOPAMAGILNP 679

RESULT 14

US-08-931-608A-5

Sequence 5, Application US/08931608A

Patent No. 6302685

GENERAL INFORMATION:

APPLICANT: lobel, Peter

APPLICANT: Smeal, David E.

TITLE OF INVENTION: NOVEL HUMAN LYSOSOMAL PROTEIN AND METHODS OF ITS USE

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: David A. Jackson, Esq.

STREET: 411 Hackensack Ave, Continental Plaza, 4th Floor

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/931,608A

FILING DATE:

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 601-1-077

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 635 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

US-08-931-608A-5

Query Match

Best Local Similarity 1.7%, Score 104.5, DB 4, Length 635;

Matches 146; Conservative 76; Mismatches 241; Indels 242; Gaps 37;

```

1 TELEFAX: (703) 413-2220
2
3 TELE: 248855 OAT UR
4
5 INFORMATION FOR SEQ ID NO: 5
6
7     SEQUENCE CHARACTERISTICS:
8
9     LENGTH: 774 amino acids
10
11     TYPE: amino acid
12
13     TOPOLOGY: linear
14
15     MOLECULE TYPE: protein

```

Search completed: March 28, 2003, 13:39:45  
Job time : 42 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2003, 13:33:34, Search time 54 Seconds  
(without alignments)  
2141.661 Million cell updates/sec

Title: US-09-769-699-2

Perfect score: 6294  
Sequence: 1 MENTOKIVTPTGPIKHYVA... DELFDLSGLPIKHCNTMEM 1203

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: PIR-73:\*

1: PIR1:\*\*\*  
2: PIR2:\*\*\*  
3: PIR3:\*\*\*  
4: PIR4:\*\*\*

Print No is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6294	100.0	1204	1	DNBE29
2	3620.5	57.5	1209	1	DNBE24
3	3583	56.9	1208	1	T42574
4	3054	48.5	1136	1	DNBEK5
5	3049	48.4	1196	1	DNBEV1
6	3047	48.4	1196	1	DNBEHF
7	3003	47.7	1197	1	A48350
8	2933	46.6	1186	1	DNBERG
9	1094.5	17.4	1375	2	J00846
10	669	10.6	1128	1	Q08547
11	633	10.1	1128	2	T43922
12	618	9.8	1145	2	S55600
13	604	9.6	1128	1	DNBEH1
14	596	9.5	1137	2	T03105
15	490.5	7.8	1132	2	T44001
16	483	7.7	1131	2	T41943
17	479.5	7.6	1160	2	A36256
18	477	7.6	1235	1	Q085M4
19	412	6.5	1191	1	A44051
20	237.5	3.8	483	2	S69894
21	137	2.2	1504	2	T117426
22	120.5	1.9	781	2	A13014
23	120.5	1.9	788	2	P98269
24	119.5	1.9	701	2	A11501
25	117.5	1.9	358	2	H83380
26	116.5	1.9	335	2	H81702
27	115.5	1.8	7576	2	T117428
28	115	1.8	823	2	B35963
29	115	1.8	2505	1	XVRRFA

30	114	1.8	1215	2	I52882
31	114	1.8	1941	2	T23979
32	114	1.8	1943	2	T23986
33	113.5	1.8	1886	2	S04921
34	113	1.8	1682	1	C70588
35	112	1.8	735	2	F87355
36	111	1.8	1699	2	T14074
37	111	1.8	3255	2	G81702
38	110.5	1.8	878	2	A87385
39	110.5	1.8	2609	2	T40399
40	110.5	1.8	1069	2	H70656
41	110	1.7	824	2	S24108
42	110	1.7	1039	2	A10122
43	109.5	1.7	977	2	C72526
44	109.5	1.7	1501	2	T45623
45	109.5	1.7	1762	2	T03222

## ALIGNMENTS

## RESULT 1

DNBE29

DNA-binding protein - human herpesvirus 3

C:Species: human herpesvirus 3, varicella-zoster virus

C>Date: 30-Sep-1988 #sequence\_revision 30-Sep-1988 #text\_change 16-Jul-1999

C:Accession: C27214

R:Davidson, A.J.; Scott, J.E.

J. Gen. Virol. 67, 1759-1816, 1986

A:Title: The complete DNA sequence of varicella-zoster virus.

A:Reference number: A27345; MUID:86306657; PMID:3018124

A:Accession: C27214

A:Molecule type: DNA

A:Residues: 1-1204 <DAV>

A:Cross-references: EMBL:X04370; NID:q59989; PION:CAA27912.1; PID:q60018

C:Genetics:

A:Gene: 29

C:Superfamily: herpesvirus DNA-binding protein

Query Match 100.0% Score 6294 DB 1: Length 1204;  
Best local similarity 100.0% Pred. No. 0;  
Matches 1203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MENTOKIVTPTGPIKHYACREVEDLDEEISFLAARSTSDALLPLMFNLTVEKFTS	60
DB	1	MENTOKIVTPTGPIKHYACREVEDLDEEISFLAARSTSDALLPLMFNLTVEKFTS	60
QY	61	SLAVVSGARTTGLAGAGITLKTTHSHFPPSVFPHGKHVLPSSAAPNLTRACNAAREPF	120
DB	61	SLAVVSGARTTGLAGAGITLKTTHSHFPPSVFPHGKHVLPSSAAPNLTRACNAAREPF	120
QY	121	GFSCQGPVAVGAVETTGAEICTPLGLEPENTILLVYATLKEKAVFMCNVPLHGGDI	180
DB	121	GFSCQGPVAVGAVETTGAEICTPLGLEPENTILLVYATLKEKAVFMCNVPLHGGDI	180
QY	181	VHINNGVIRIPFPVQLEFMDVNRVLPDPNTTHRSIGEGVYPTPTPYNTGLCHLHDC	240
DB	181	VHINNGVIRIPFPVQLEFMDVNRVLPDPNTTHRSIGEGVYPTPTPYNTGLCHLHDC	240
QY	241	VIAPMVAVLRVNRVAVARGAHLAFDENHGAVALPPIITYTPSSSGTTTACGARN	300
DB	241	VIAPMVAVLRVNRVAVARGAHLAFDENHGAVALPPIITYTPSSSGTTTACGARN	300
QY	301	DVNSTSKPSPGGEFERRIASMAADTALHAIEYNTGTIETPTDIKEMPFIGNETLP	360
DB	301	DVNSTSKPSPGGEFERRIASMAADTALHAIEYNTGTIETPTDIKEMPFIGNETLP	360
QY	361	RLNLGSTARVAVGIVGAMVSPNSALYTEVEDSGMTEADGGGSPFNFFYQFAGHL	420
DB	361	RLNLGSTARVAVGIVGAMVSPNSALYTEVEDSGMTEADGGGSPFNFFYQFAGHL	420
QY	421	AANDOTDRDCHVLSSTSGSSNTFSVDYLAICGFGAPLALRLILFYLRDCAAGATGCH	480

```

10 421 AANPQTDRIKHWISSQSSNTFESVIALICFGAPLAPLLEYLERTCAAGTGGH 480
11 481 GIALKAVYTCIPSEISLCEKHTIRVCAHTTVHRIKQMPHKGQATKQIPGVGIMNSQ 540
12 481 GIALKAVYTCIPSEISLCEKHTIRVCAHTTVHRIKQMPHKGQATKQIPGVGIMNSQ 540
13 541 YSNKDPGLNVAAYLLIRKGDGTEAAKATIMQDTRATLERLFDLEQEDLFGAPGCSSE 600
14 541 YSNKDPGLNVAAYLLIRKGDGTEAAKATIMQDTRATLERLFDLEQEDLFGAPGCSSE 600
15 541 YSNKDPGLNVAAYLLIRKGDGTEAAKATIMQDTRATLERLFDLEQEDLFGAPGCSSE 600
16 601 GLSSVIVDHPTRIRILIDTLKARIEDDTTQPMKVLVETRDYKIREGSEATHSNALTDEPY 660
17 601 GLSSVIVDHPTRIRILIDTLKARIEDDTTQPMKVLVETRDYKIREGSEATHSNALTDEPY 660
18 601 GLSSVIVDHPTRIRILIDTLKARIEDDTTQPMKVLVETRDYKIREGSEATHSNALTDEPY 660
19 661 SCAPGPTNFEVAKRIHAAVVDIALISQCHGVYGVGVGEGKRNKRPVUVAHKKHVLDFMG 720
20 661 SCAPGPTNFEVAKRIHAAVVDIALISQCHGVYGVGVGEGKRNKRPVUVAHKKHVLDFMG 720
21 721 GFTSTRITVTLSSEHVSAPNPTLQDAPAGRTFDGDLARVSEVIRDIRVNRVFSGN 780
22 721 GFTSTRITVTLSSEHVSAPNPTLQDAPAGRTFDGDLARVSEVIRDIRVNRVFSGN 780
23 781 CTNLSSEAAAKRVGLASAVQROEKRYDMHGLGELKQFIRGLLPPRGMPNPKSPNPM 840
24 781 CTNLSSEAAAKRVGLASAVQROEKRYDMHGLGELKQFIRGLLPPRGMPNPKSPNPM 840
25 841 FWTLLQNRNMPAKLHEEELIAAKKRELEHYAALNINLPPICIGELADPYMANILIK 900
26 841 FWTLLQNRNMPAKLHEEELIAAKKRELEHYAALNINLPPICIGELADPYMANILIK 900
27 901 YGDHSQYLINTTSTTTAAPPDPSSVLMHMKPDVTSADTETQAKALLEKTENIPELM 960
28 901 YGDHSQYLINTTSTTTAAPPDPSSVLMHMKPDVTSADTETQAKALLEKTENIPELM 960
29 961 TTAFTSTHIVRAAMNRPVAVLGISIKYHGAAGNNRVQANMSTLNGKKNVCPLETFD 1020
30 961 TTAFTSTHIVRAAMNRPVAVLGISIKYHGAAGNNRVQANMSTLNGKKNVCPLETFD 1020
31 1021 KTRKRFIACPRGCGTCPTVGPSSGNKRETTLSQVKGCIYVSGAMVOLAIVATVRAVAGAR 1080
32 1021 KTRKRFIACPRGCGTCPTVGPSSGNKRETTLSQVKGCIYVSGAMVOLAIVATVRAVAGAR 1080
33 1081 AOHMAFDMHSLIDBEPLAKDLLEHLDJLOTLFTWIVGCALEAVKILDEKTTAGIDCEP 1140
34 1081 AOHMAFDMHSLIDBEPLAKDLLEHLDJLOTLFTWIVGCALEAVKILDEKTTAGIDCEP 1140
35 1141 PTNLAFNEUSCEPSHITTSNVLNTSGNSISGSTVPELKRPEDELEFDSGIPKIGNIT 1200
36 1141 PTNLAFNEUSCEPSHITTSNVLNTSGNSISGSTVPELKRPEDELEFDSGIPKIGNIT 1200
37 1201 MEM 1203
38 1201 MEM 1203
39 1201 MEM 1203
40 1201 MEM 1203
41 1201 MEM 1203
42 1201 MEM 1203
43 1201 MEM 1203
44 1201 MEM 1203
45 1201 MEM 1203
46 1201 MEM 1203
47 1201 MEM 1203
48 1201 MEM 1203
49 1201 MEM 1203
50 1201 MEM 1203
51 1201 MEM 1203
52 1201 MEM 1203
53 1201 MEM 1203
54 1201 MEM 1203
55 1201 MEM 1203
56 1201 MEM 1203
57 1201 MEM 1203
58 1201 MEM 1203
59 1201 MEM 1203
60 1201 MEM 1203
61 1201 MEM 1203
62 1201 MEM 1203
63 1201 MEM 1203
64 1201 MEM 1203
65 1201 MEM 1203
66 1201 MEM 1203
67 1201 MEM 1203
68 1201 MEM 1203
69 1201 MEM 1203
70 1201 MEM 1203
71 1201 MEM 1203
72 1201 MEM 1203
73 1201 MEM 1203
74 1201 MEM 1203
75 1201 MEM 1203
76 1201 MEM 1203
77 1201 MEM 1203
78 1201 MEM 1203
79 1201 MEM 1203
80 1201 MEM 1203
81 1201 MEM 1203
82 1201 MEM 1203
83 1201 MEM 1203
84 1201 MEM 1203
85 1201 MEM 1203
86 1201 MEM 1203
87 1201 MEM 1203
88 1201 MEM 1203
89 1201 MEM 1203
90 1201 MEM 1203
91 1201 MEM 1203
92 1201 MEM 1203
93 1201 MEM 1203
94 1201 MEM 1203
95 1201 MEM 1203
96 1201 MEM 1203
97 1201 MEM 1203
98 1201 MEM 1203
99 1201 MEM 1203
100 1201 MEM 1203

```

## RESULT 2

DNA binding protein - equine herpesvirus 1 (strain Ab4p)

C:Species: equine herpesvirus 1

A:Note: host Equus caballus (domestic horse)

C:Date: 10 Sep-1992 #sequence revision 30-Sep-1992 #text-change 16-Jul-1999

C:Accession: E16798

Submitted to GenBank, March 1992

A:Description: The DNA sequence of equine herpesvirus-1.

A:Reference number: A16805

A:Accession: E16798

A:Molecule type: DNA

A:Accession: 1209 <HEP>

A:Accession: 1209 <HEP> NID:9440791; P108:AAH02466.1; P1D:9330823

A:Accession: 1209 <HEP> NID:9440791; P108:AAH02466.1; P1D:9330823

A:Accession: 1209 <HEP> NID:9440791; P108:AAH02466.1; P1D:9330823

A:Accession: 1209 <HEP> NID:9440791; P108:AAH02466.1; P1D:9330823

A:Accession: 1209 <HEP> NID:9440791; P108:AAH02466.1; P1D:9330823

A:Accession: 1209 <HEP> NID:9440791; P108:AAH02466.1; P1D:9330823

A:Accession: 1209 <HEP> NID:9440791; P108:AAH02466.1; P1D:9330823

A:Title: The DNA sequence of equine herpesvirus-1.  
A:Reference number: A16813; MID:9235566; FMD:1318606  
A:Contents: annotation, possible protein-coding frames  
A:Note: neither amino acid nor nucleotide sequence is given  
C:Genetics:  
A:Gene: 31  
C:Superfamily: herpesvirus DNA-binding protein  
C:Keywords: DNA binding

Query Match 57.5%; Score 3620.5; DR 1; Length 1209;

Best local Similarity 57.1%; Pred. No. 1.5e-266;

Matches 698; Conservative 191; Mismatches 299; Indels 35; Gaps 14;

```

11 MENTQTVVPTPTPLGYVACRVEDLDEISFLARSTSDIALPLMKNLVEKTFPS 60
12 1 MESAPKTVSLPSPSLQVYVAKKASIVGIVSIITAKSVSDIALVAVLKGILVDFV 60
13 61 SLAVSAGPTTGLAGATLTKLTTSHFYSVFEHAKHVI PSSAAPNLTPAANAEERF 120
14 61 NNAVVAISKTTGSLGHTGLKLTSPHFNPNAEFVFGSVIGASSNPNLTPACEAARPR 120
15 121 GFSRCGCPVDDGAVETGACIEICRIGLFPFNTLIYVIALHKAAPKCNVPLHGGCIDI 180
16 121 GESAESPPVENAVETSGEICASLNLSPETALYLVTSEFEKMYVONTFLHKGSTST 180
17 181 VHNHGDVRIPLFPVQLTMPDVNRLVPPDFTNHSIGEGEYVPTPEYNTGLCHLIDG 240
18 181 VLDQDQAKPIPIYVQVLMPOVNRILASEPNAKHSIGEFVYSNPNSDLCRLHGY 240
19 241 VIAPPAVALKRVNRTAVAGAAHLADENHEGAVLPDITITYFQSS--SSGTTAKGAR 298
20 241 VGPAAVALKRVNRLDVGAAHALLADENHEGAVLPDITITYFQSS--SSGTTAKGAR 300
21 299 RNDVNSTKSPSGCFERRLASIMADTALHAVERFNTGYEFTPDITKHWMFQMGCH 358
22 301 QED-GSGSNASSSGTERELASVMAADTALSDVISMGAGITDELPSVEDMPLSSGDT 359
23 359 LPRNLALSYTAKVAVIGAMVFPNSALYLTEVEDSGMTEAKDGGPSPFNRYFQAP 418
24 360 -BSIFALGAYAAARLSGLVAGAMVPSANSVLMFVVDGCPAIDKDS-NHSYHRYLAAV 417
25 419 HIAANPDTPDGHV-----SSQTSSTSTSPSVYALTFYFCAPLAPLIFYEPD 472
26 418 IYAGNPQIDKDGVLPHITADQVAPLNGSNQEPSTLDYLALAGCPQILARLLEYLERD 477
27 473 AGAFTG-GHGDALKYVTGTFDSRIPCSLCEKHTRPVCAHTVARIKQMPHKGQATKQIP 531
28 478 AGTEGGRNFTDALRYLANTLESQVPGCLNQATRPACHTTLHRLQRLPRGAPVRAPI 537
29 532 GYFGTNSQYSDQDPLGANTAPYLILKPRGQVTEAAKATIMQDTRATLEKFLIDLEDERLL 591
30 538 GILGIMNSAVSDQDVIGNAYGALAKRPND-NIAPKSIIMQDTRATLEKFLIDLEDERLL 596
31 592 DR-----GAPCSSEGLSSVIVHPIFKRILIDTLKAREDTTQPMKVLVETRDYKIREG 646
32 597 DKETLAQASPCAP--TSVYHDOASTIGLLSNKIDTEGAAGDFMTLYEARDPKIREBL 654
33 647 SEATHSMALTPDPYSCAFCEPINFVAKRTHLAAYVDIALISQCHGVYGVGVGEGKRNKRP 706
34 655 ADANHTMISLIDPYSFSPFVTSFLARTRFAVLQDLVLSQCHLEFYGVSGVERNRKNGF 714
35 707 QVLLRRRFVDLNGGISTIKSTITVLSSEGPVAPNPTLQDAPAGRTFDGDLARVSEV 766
36 715 QVLLRRRFVDLNGGISTIKSTITVLSSEGPVAPNPTLQDAPAGRTFDGDLARVSEV 774
37 767 PDIPKVPVVFESGNPTNLSEAPAPLVGLASAVQROEKRYDMHGLGELKQFIRGLLPP 826
38 775 RDLRYKKNRVLNFGANGKSEARARVAGASAVRPDRKSNLNLNAGVGLVQGYRVLFP 834
39 827 KQMPNKSPPNPMQVWTLLQNRNMPADKLTREHITTTAAVKRPLEEYAAININLPPIC 886
40 835 RCHPQIDTPNPQVFWTLQNRNMPARLLEKSDIETITAIKRFSEDEYSAININLPPNI 894

```





A:Reference number: A28601; MUID:88179536; PMID:2833010  
 A:Accession: A28601  
 A:Molecule type: DNA  
 A:Residues: 1-1196 <MG>  
 A:Cross references: GB:M20165; NID:q330120; PID:AAA5793.1; PID:q330121  
 C:Genetics:  
 A:Map position: 0.38-0.409  
 C:Superfamily: herpesvirus DNA-binding protein  
 C:Keywords: DNA binding

Query Match 48.5%; Score 3054; DB 1; Length 1196;  
 Host Local Similarity 49.8%; Pred. No. 1.8e-223;  
 Matches 609; Conservative 200; Mismatches 366; Indels 48; Gaps 17;

1 MENTOKTAT---VTPGPGYVY--ACRVADLDEISFLAARSTSDALPLMNTIVE 55  
 116 ARRRGFSKCGPPVAVAVETTGAEICTRLGLEPENTILYLVVTLFKEAVMCAVFLHY 175  
 119 ARRRGFSUYTPRPDGLKHETTGALCEKLDIDPRKALLYLAVTGPKRAVCIINNTFLH 178

176 GGLDVIHINHDVIRPLEPVOLPEPDVNRLLVDPFNTIHRSIGEGFYPPPEYNTGLCH 235  
 179 GGSOKVTITGAAVHRIPIYPLDLPDPSKVIAEPFNANHSIGENFTYPLPFRKPLNR 238

216 LHHKVALPMAVALKRVNVAVARCAALAHDEHNECAVLDPPDIYTYFOSSSCHTTAR 295  
 219 LTFEVAVPAAVAVLEKRVNDAVARAAHLDADENIEGALLADITFEAASQG--KTPR 296

296 GARRDNVSTSKPSGGEEERLASIMAADTALHAEVIFNGGIVETPEPDKEMPMFGM 355  
 297 GGR-----DGGCKGACGPFORIASVMAGDAAALHESIVSAVAPDEPPLDLSAMFLCEGQ 351

456 ECLTPELNLALGYSYTAARVAVIGAVFSPNSALYLTLEVEDSGTEAKDGSPGSENFYQF 415  
 452 DTAARAAVAVAYLAKAGLWAMVSTNSALHLEVEDDAGFADPKHSK--PSFYREFLY 410

416 AGPLAANVQIDIRKCHV-----SSQSTGSSNTPEFSVITALLCGCAPILAMLLTYL 468  
 411 PGTIVAAVNPVDREHVAVGFEGREPTALVAGTQ--EFAEHLAMLCGSPALLAMLEYL 469

469 EKDCAAGTGGHG--DALKYVTGTEDSEIPCSLCEKHETPVCANTVHRLRORPFGQAT 527  
 470 EKKCKGVAVCGQEMHGVFRVADSNQTDVPCNLCPTDTHACVHTILMKLRANHKPKASAA 529

528 KQPIGVETSTMSYSYQVOPVGNVAYLLFLRPGLDTEAKATMDQTYATLEPFLDLEQ 587  
 540 KGAIGVEGTMSYSDIVLGNVAFSAIKR--ADGSEFARTIMQETVRAATERVAAELET 588

588 EKLIDKRAVCSSEGISVIVDHPFRRLIDPLMAHLEDTOPMVIVETNHYKLRHGLS 647  
 589 LQYVQAVPTAMKLETTITNREALHTVANNVROVDEVEDLMNLEVGGRKFRDGG 648

648 EATHSMALTFDPSYSGAFPTNLFVKPETHLAVVQPLALSYTHVVFYVQVVEYENFRNFGU 707  
 649 EAHNAHSTLIDPYACGPDILQILGRSNLAIVQDIALISQCHGVAGGVSQGRNRNFO 708

708 PVLRRRVDELNGCFITSPSTITVLTSE--PVSAPNPPLTGQDAPAGPTDGLARVSEVI 766  
 709 PVLRRRVDEMNNQELSKRTITVALISFGMAICAPSLTAGQTAAPSSSEGGVARTLTGFP 768

767 RDIIVKNNVVSQNCNTINSEARARLVGLASAYOQKRVMLHCAIGFLIKQFGILPP 826  
 769 KELIVKSVRLTAGSANAASEAKARVASLSQAYQKPKRVLDLLGSLGFLKQFIATLFP 828

827 PGMPPNSKSPDQFWFLLQGNQMPADKLTHEETITIAAVRFTEEVAAINFINLPTCI 886  
 829 NGKIPGNSQNPQWFWALQNNQIPARLISRDDIETIAFKKFSIDYCAINFINLANNV 888

QY 887 GELAQFYMANLLKCYDHSQYLINILITSLITGAKKRPDSSVLHMLKIVLSAADIETVA 946  
 DB 889 SELAMTYMANQLIRYCDHSTYFINTILITLITAGSRPPSYQAAMV--SAGGASILEGA 945

QY 947 KALLKTEENLPELMTTAPSTLIVRAAMQRPVVYIGISISYHGAAGNNRPQAGNMSG 1006  
 DB 946 RALMDAVDAHPPAMTSMFASCLLRLPYMAARVYVIGLISISKYGMACNDRPQAGNMS 1005

QY 1007 LMGKNVCPLEFPDTPRFTIACPPRGFTCPVTGPPSSGNRETTLSQVYGVITVSQAVY 1066  
 DB 1006 LMGGNACPLLEFDTRKRVLAACPRAGFCVCAASNLGGAHSSLCRQLHGLITSEQAAVA 1065

QY 1067 LAIVATVAVAGARQAHAFDMLSTIDDEPLARDILEHDLQIQLTETPWVVEKAL--- 1123  
 DB 1066 SSVFAITYKSLSPRTQVULIEDMLALLEDEYLSSEKMETLAKALEKNGEMSTDALLEYA 1125

QY 1124 -EAVKILDEKTIACGDEPTNLAFNPD--SCPSHDTSNVINISGSNLSGSTVPLQIKRP 1180  
 DB 1126 HEAFALVSQGNAGE-----VFNPGRGGR-----IDNATPPGQVCAVAPAGRKRA 1173

QY 1181 PEDDELFDLSGIPIRKGNITTEM 1203  
 DB 1174 FRGIDPFG-KSPDPRKKGLFLDM 1195

## RESULT 5

DNBEV1  
 major DNA-binding protein UL29 - human herpesvirus 1 (strain 17)

C:Species: human herpesvirus 1  
 A:Note: host Homo sapiens (man)

C:date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 16-Jun-2000  
 C:Accession: A03790; B30085

R:Quinn, J.P.; McGeoch, D.T.  
 Nucleic Acids Res. 13, 8143-8163, 1985

A:Title: DNA sequence of the region in the genome of herpes simplex virus type 1 cont

A:Reference number: A93601; MUID:86067223; PMID:2993714

A:Accession: A03790

A:Molecule type: DNA

A:Cross references: GB:X03181; GB:M12356; NID:q59862; PID:CAA6640.1; PID:q59863

A:Experimental source: strain 17

J:McGeoch, D.J.; Dalrymple, M.A.; Davison, A.J.; Dolan, A.; Frame, M.C.; McNab, D., P  
 J:Gen. Virol. 69, 1531-1574, 1988

A:Title: The complete DNA sequence of the long unique region in the genome of herpes

A:Reference number: A30083; MUID:88274327; PMID:2839594

A:Accession: B30085

A:status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1196 <MG>

A:Cross references: GB:D10879; NID:q221721; PID:BAAN01675.1; PID:q221720; GB:D00417

C:Genetics:  
 A:Gene: UL29

A:Map position: 0.38-0.409

C:Superfamily: herpesvirus DNA-binding protein  
 C:Keywords: DNA binding

Query Match

48.4%; Score 3049; DB 1; Length 1196;  
 Best Local Similarity 49.8%; Pred. No. 4.4e-223;

Matches 609; Conservative 199; Mismatches 367; Indels 48; Gaps 17;

QY 1 MENTOKTAT---VTPGPGYVY--ACRVADLDEISFLAARSTSDALPLMNTIVE 55  
 DB 1 METKRTATITKVPGLGTYVARACPSGEGL--LALLSARSGSDAVAVPLVGLTVE 58

QY 56 KTFSSLAVSGARTTGLAGAGITLKLTHSHYPSVFPFGCKHVLDPSSAAILNRACNA 115  
 DB 59 SGEFANVAVVGSRTTGLGTAVALSLKLTPSHYSSVYFPHGCKHLPSTQANPLRLCKR 118

QY 116 ARRRGFSKCGPPVAVAVETTGAEICTRLGLEPENTILYLVVTLFKEAVMCAVFLHY 175  
 DB 119 ARRRGFSUYTPRPDGLKHETTGALCEKLDIDPRKALLYLAVTGPKRAVCIINNTFLH 178

```

QY 176 GOLDIVHINHGVIPIPLFVYVCLFMDVNR.LVDPFNTNHRSLGEGFVYPTPEYNTGLCH 235
DB 179 GCGDAVTTIGCAVHNRIPVYPIQI.FMPDRSRVLAEPNNAHRSIGEFYTPLEFPNPNLNR 238
QY 236 L.IHDCVIAPMVAALRVNVTAVARCAAH.APDHNBGAVALPDDITYTYVOSSSGCTTTAR 295
DB 239 LLEFVAVPAAVALR.RVNDVAPAAHAFDENHEGALPADIFTEAEASOG--KTRR 296
QY 296 GARRNDVNSTKPSPCGERRIASIMADPTALHAVEITNCTGEPTEPTIKEMPFICM 355
DB 297 GCR-----DGGCKGAAGCEQRLASVMAQDALALEISVMAVFDEPPDISAMPLEBO 351
QY 356 EGTLEPLNALGSYIARVAVIGAMVSPNSALYTEVEDSGMTEAKDGKGPSENFYOF 415
DB 352 DTAARAANAAGAYLARAAGTICAMVSTNSALHLETVDAGPADPDHKK--PSYAFPIV 410
QY 416 AGPHLAANQOTDRIGHVL-----SSOSTGSSNTERSVDYIALICGKAPILARILLYL 468
DB 411 PGIHVAANQVIRESHVYVGFEGRETPALVGTU--EFAGEHLAMLCGFSPALAKMLFYL 469
QY 469 EPCFAGAFTHG--CAI KVTCTFSEISLCEKHTPRVCAHTTVHRLKRMHPPGQAI 527
DB 470 KRCDAVAVIGREMOVFRVVAUSNOTDVPCNICTFDRACVHTTLMKRAHPRKASAA 529
QY 528 KQFVGFETMNSVYSDCEFLGNVAPYLLIRKPDQTEAKATMODTYRATLERLEFLDQ 587
DB 530 KQALGVFETMNSVYSDCEFLGNVAPYLLIRKPDQTEAKATMODTYRATLERLEFLDQ 588
QY 588 KRLIRKAPCSSEFGLSSV.VIHPHPRKILDTPLARILEOTTPQPMKVIATVTRYKI.PEGIS 647
DB 589 LQYVQAVPTAMGRLETTITNEALHTVYNNV.KUYVREVEJLRVLVGRNKFEDSLA 648
QY 648 EATHSMALTFEDPSGAFCTPITNFKRLAVVQDIALSQCCHVFYQVVEGKRNFRQ 707
DB 649 EANHMSLITIDYACGCPCLDQLGRSNLAVYODIALSQCCHVFYQVVEGKRNFRQ 708
QY 708 PYLRRFVDLEFGFISTSTIVTSEG--PVSAFNPITLQDAPAGTPEGDLARVSEVI 766
DB 709 PYLRRVMDMFNGFLSAKTLTVAL.SEGAAICAPSLITAGOTAPAESSEFGDAVAVTLGFP 768
QY 767 PHLVKNVYVSGNCTNI.SFAAPARI.VGIAVYQROEKFVMDLHGAICPLKQFHGLP 826
DB 769 KELVAKSPFLFNGASANA.SEAKAPVASTYAKRDKRVDDLGLPGLLQFHAIIFP 828
QY 827 KCMPTNSSTNQWFWLILQNMOMADKLTHERKITIAAVKRTFEYAAINFLPPTCI 886
DB 829 NCKPQSNQPNQWFWLILQNMOMADKLTHERKITIAAVKRTFEYAAINFLPPTCI 888
QY 887 GELAGFYMANLTKYCDHSOYLINTLSTIITGARRPPDSSVLMWIRKQVTSADLETQA 946
DB 889 SHLAAYTMANOILTKYCDHSOYLINTLSTIITGARRPPDSSVLMWIRKQVTSADLETQA 945
QY 947 KALIKRTNIPPLMTAFTSTHIVRAANQORPMVYLGISISKYHGAAGNNRVFOAGNMG 1006-
DB 946 PALMDAVVAHPAMTSMFASCNLDLPVMAAPVYVGLSTISVYCAKAGNDPVGQANMA 1005
QY 1007 LMGCAVNCGLTFEDTKRRITACPRGCGTGVYIGSSGRRFTTISQVNGTIVSGCAMA 1066
DB 1006 LMGCAVNCGLTFEDTKRRITACPRGCGTGVYIGSSGRRFTTISQVNGTIVSGCAMA 1065
QY 1067 LAIVTVAVAGARQIHMFDMLSTLDEFLAROLEELHDOJLITLFWPTVEAL--- 1123
DB 1066 SSVFAVAVYSLGPKVUQJLHLMIALHDEYLSERKMLTARALEGNOMKSDAALHVA 1125
QY 1124 -EAAVILDEKTTAGDGETTINLAFNFD--SCPSHDUTTSNVLNISGSI.SGSTVPIKRP 1180
DB 1126 HEAEVLEVSQGNAGE-----VFNEGDFGE-----DONATFEQSGAGVAFAGRKRA 1173
QY 1181 PEUDPLPLDLSGLPIKHQNTIMEM 1203
DB 1174 FHGDPPFC--EGPPDKKGLDILDM 1195

```

```

RESULT 6
DNEHF
DNA-binding protein - human herpesvirus 1 (strain F)
C:Species: human herpesvirus 1
A:Note: host Homo sapiens (man)
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #ext-change 07-Jun-1996
C:Accession: D29242
R:Hammerstrand, W.; Conraths, F.; Mankertz, J.; Pauli, G.; Ludwig, H.; Buhk, H.J.
Virology 165, 388-405, 1988
A:Title: Conservation of a gene cluster including glycoprotein B in bovine herpesvirus
A:Reference number: A94381; M01D:85306231; PMID:2841793
A:Molecule type: DNA
A:Residues: 1-1196 <HAM>
A:Cross-references: GB:M21629
C:Superfamily: herpesvirus DNA-binding protein
C:Keywords: DNA binding

```

```

Query Match          48.4%; Score 3047; DB 1; Length 1196;
Best local similarity 49.8%; Pred. No. 5,3e-223;
Matches 609; Conservative 199; Mismatches 367; Indels 48; Gaps 17;

```

```

QY 1 MENTQTVT---VPTGPIGVY--ACRVEDLDEISFLAARSTSDIALPLMRNITVH 55
DB 1 METPKTATTIKVPPGLGVYARACPSSEGL--LALLSARSGSDVAVAPLVGLIVE 58
QY 56 KPTSSLAVVGARTGLAGAGITLKTSHFVSVFVPHGKHVLPSSAAPNITRACNA 115
DB 59 SCFPAVAVVCSPTTICLTCTANSLKTTSHVSSVYVPHGPHLNDSTQAPNITRICK 118
QY 116 ABERFGRSQGPPVYVAVETTGAEICTRLGLEPENTILYLVLTALKEAVFMQNFY 175
DB 119 AKNHGPKSYIYPRPDGLKHEETGEALCERGLDPRALITLVYVEGFKKAVSINNPLHL 178
QY 176 GOLDIVHINHGVIPIPLFVYVCLFMDVNR.LVDPFNTNHRSLGEGFVYPTPEYNTGLCH 235
DB 179 GCGDAVTTIGCAVHNRIPVYPIQI.FMPDRSRVLAEPNNAHRSIGEFYTPLEFPNPNLNR 238
QY 236 L.IHDCVIAPMVAALRVNVTAVARCAAH.APDHNBGAVALPDDITYTYVOSSSGCTTTAR 295
DB 239 LLEFVAVPAAVALR.RVNDVAPAAHAFDENHEGALPADIFTEAEASOG--KTRR 296
QY 296 GARRNDVNSTKPSPCGERRIASIMADPTALHAVEITNCTGEPTEPTIKEMPFICM 355
DB 297 GCR-----DGGCKGAAGCEQRLASVMAQDALALEISVMAVFDEPPDISAMPLEBO 351
QY 356 EGTLEPLNALGSYIARVAVIGAMVSPNSALYTEVEDSGMTEAKDGKGPSENFYOF 415
DB 352 DTAARAANAAGAYLARAAGTICAMVSTNSALHLETVDAGPADPDHKK--PSYAFPIV 410
QY 416 AGPHLAANQOTDRIGHVL-----SSOSTGSSNTERSVDYIALICGKAPILARILLYL 468
DB 411 PGIHVAANQVIRESHVYVGFEGRETPALVGTU--EFAGEHLAMLCGFSPALAKMLFYL 469
QY 469 EPCFAGAFTHG--CAI KVTCTFSEISLCEKHTPRVCAHTTVHRLKRMHPPGQAI 527
DB 470 KRCDAVAVIGREMOVFRVVAUSNOTDVPCNICTFDRACVHTTLMKRAHPRKASAA 529
QY 528 KQFVGFETMNSVYSDCEFLGNVAPYLLIRKPDQTEAKATMODTYRATLERLEFLDQ 587
DB 530 KQALGVFETMNSVYSDCEFLGNVAPYLLIRKPDQTEAKATMODTYRATLERLEFLDQ 588
QY 588 KRLIRKAPCSSEFGLSSV.VIHPHPRKILDTPLARILEOTTPQPMKVIATVTRYKI.PEGIS 647
DB 589 LQYVQAVPTAMGRLETTITNEALHTVYNNV.KUYVREVEJLRVLVGRNKFEDSLA 648
QY 648 EATHSMALTFEDPSGAFCTPITNFKRLAVVQDIALSQCCHVFYQVVEGKRNFRQ 707
DB 649 EANHMSLITIDYACGCPCLDQLGRSNLAVYODIALSQCCHVFYQVVEGKRNFRQ 708
QY 708 PYLRRFVDLEFGFISTSTIVTSEG--PVSAFNPITLQDAPAGTPEGDLARVSEVI 766
DB 709 PYLRRVMDMFNGFLSAKTLTVAL.SEGAAICAPSLITAGOTAPAESSEFGDAVAVTLGFP 768

```



770 BYKABYTESCNGENT SEABADIVCI KAVODORBYDUF IGAT C

QY 1128 ILDEKTTAG---DGETPTNLAFNFDSC---EP SHDTTSNVLNISGSNISGSTVPGI,KRPP 118



```

Db 18 PVGACGYIVYVPEKCEPQEFASLIGNRVGADAMSLPLLSGLTVRANSPVNAKAVHKID 77
QY 71 TGLAGAGITLLITLTSHEPVSFVFGKHVLPSSAAPMLTRACNAAREGFSRCQCPV 130
Db 78 M-----TTLSSVRSANREAIIVEFNTEPTTFEGPGLDLSGSDARMLFGYTSPT-PRH 130
QY 131 DGA-----VETTGAEICRIGLEPENTILYVLTALFKEAVPMCNVFLHYCGDILVHINHG 186
Db 131 DLKDLIDIKDLAFETYK-----DSCEMSVVTSEKRELFENGLVPLISGQOKVJNGR 185
QY 187 DVIRIPFPVLPMPDNKRLVDPDPNTHRSIGGEVYPTPPYNTGCLHLHDCVAPMA 246
Db 186 EAVKIPLYDEDLFSKSHHL-----PRFYIPSVSKYLDLSLFTSIA 226
QY 247 VALRVNRTAVARGAHLAFDENEGAVLPDITYTFQSSSSGTTARGRANDVNSTS 306
Db 227 QALRIREVESIRALEQS IHDOYKIAKV-----VNSK- 259
QY 307 KPSPSGCHERLASIMADTA-----LHAEVIFNTGI-YEETP-----TDIKEMPMF 352
Db 260 -----EFALQAVKQDQSAFVYIDCIAMELAISYGLSFLFEPQECALLDYTSWEIF 311
QY 353 IGMEGLPRLNALGYSYTAARVAVIGAVVESPNALYLTEVEDSGMTEKDGPGPSFNRF 412
Db 312 DITTEGGRITKAIQDMAMMSVHYTHLFTNSVLYLTKINKOTQTNKSDN---LYNSY 368
QY 413 YQFAGPPLAANPOTDRDGHVLSQSTGSSNTEFSVDYALICGAPLLARLLELCRD 472
Db 369 FMOHGLSYAAEATQKEMDPAFSGAVKFGNGSYTLFHLALASSPSPHLLANCGYMQDCQ 428
QY 473 AGAPTOGHGALAKVYCTPSEIPCSICEKHTRPVCAHTYHRLRQMPRGQATROPIC 532
Db 429 HOKSTTNSNTSVMOYVTVAVSELCELCQKCPACGHTTFLRLKDEPRVLLSSQRRDPY 488
QY 533 VESIMNSYSDCDPLGNVAYLILKFGD-----QTEAAKAIMDYATLIERLIDLEJE 588
Db 489 VAVGVSQVYNDLMLGNFAF---REKEDDMVONTBECEKTYMOLIONVEKL----- 538
QY 589 RLDDKAPGCSSEI---SSVVDHPTERRILDTLARIEDOTTQPMKVLVETRDYKIREGL 646
Db 539 -----ATIGITEGLSGDLITDIOSEFLKREIDNVVDNEVIKFINCLV-NNINFETI 592
QY 647 SEATSMALTFDYSACGPTNPLVYKRTHLAVYODLALSGCHVYVYQ----- 695
Db 593 KGVHIVLHGCCNFMQACMFLNLFKSVLAIIDICLV-AMITEODNPSGMMPE 650
QY 696 -----QVBSRNFN-----QFQVLRFRFYDLENGEFTSTRSITVLTSEGPVS 738
Db 651 WIKMHYQTIWNTNFKSSCIDKGVLTGAHKKVYHGDMPCDPINVD-----S 694
QY 739 APPTIGODAPARTDGLARVSEVIRIDIRKKNRVESGNTNLSEAPARLVGLASA 798
Db 695 ALN---GQIVPV---KMOVRLKALITVPTKIKIKNRIVFSN---SMTEALIOAGFI----- 742
QY 799 YOROEKRVYD--MLHGAIGLILKQFHGLFPRGMPNPKSPQMFV--TILORONMPA-DK 854
Db 743 --SSSTKQSDYITGYMKFLNSLHKVIFP-----DAKSALYMHTEFSQKQIPLRG 794
QY 855 LTHEETTTA-AKRTTEEAALNFINLPPTCIGELAQYMANLILKYCHSYYLNTL- 912
Db 795 ISKENIIEIANYIEAGSKMDMNVDIIPPTLLAAKAVRLNMTLFTFCOTOFAYATLQ 854
QY 913 -----TSIITGARPRDPSVLHMKRDVTSADIEQAK---ALLEKTENLELMWTA 964
Db 855 CLLPLOVYSAITEPH---VLH---QOSTISVDYDLSISINKQALIVOTLLKEDIATIG- 907
QY 965 TSHIVRAAMNORPVVGLISIKYHCAAGNNVFOAGN---WSLGNCKNVCPLFT-PD 1020
Db 908 -----KHPIVTVPLVYVKYTGINGNTQIFQCGNLGYFMGRGVDRNLLPDSGGR 957
QY 1021 RRRRFITACPRGEGICPVG---FSSGNRET-----TISDVYRGILVSSGAM 1064
Db 958 KONSATVMMKRRHIFMPPIVANIILKRTSNINLTFEVEETIRKNVOTIIPDKDNINIPDNV 1017

```

```

QY 1065 VOLAIATVAVRAVARGAQMAMAFDMLSLTDEFLARDLELHDOILLOTLPFWEGALIE 1124
Db 1018 IEL-----VKIGSECEMTEDEDDQETLYGDIYINSDELMSWSPULLTDSGTMSVE---S 1068
QY 1125 AVKILDEKTTACDGTGPTNPLAR 1146
Db 1069 VTKILGSRK-----QEDCNLEP 1085

RESULT 12
S55600
single-stranded DNA binding protein 06 - equine herpesvirus 2
C:Species: equine herpesvirus 2
C:Date: 27-Oct-1995 #sequence revision 03-Nov-1995 #text change 26-Aug-1999
C:Accession: S55600
R:Telord, E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.
J. Mol. Biol. 249, 520-528, 1995
A:Title: The DNA sequence of equine herpesvirus 2.
A:Reference number: S55594; MUID:95302501; PMID:7783207
A:Accession: S55600
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1145 <TEL>
A:Cross-references: GR:020824; NID:q695172; PIDN:AAC13793.1; PID:q695178
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995
C:Superfamily: herpesvirus DNA-binding protein

Query Match 9.88; Score 618; DB 2; Length 1145;
Best local similarity 23.18; Pred. No. 4e-38;
Matches 295; Conservative 182; Mismatches 540; Indels 260; Gaps 47;

QY 4 TOKTVYVTPGLGYVYACRVEDDLDELSFLAARSTSDLLPLMRULYEKTTSSLA 63
Db 17 TOASY-----GFCGYVLYLPDTPPEVEASLIGNHAGGEVSLPLLSGLTVADDFHNVK 72
QY 64 VSGARTTGLAGAGITLTKLTSHEPVSFVFGKHVLPSSAAPMLTRACNAAREGFS 123
Db 73 AVHKRLDPA-----TVSKASAVHREIVTAVANACFPFTAPGRLBBLVANSRLRGYA 126
QY 124 RCGPVPDGAIVETGA-----EICTRLGLEP-ENTILYVLTALFKAVTMCNVEFLHG 176
Db 127 EPE-----ERAGGAARPELADLGLHLPQASSHLAGVAVTSEFKERLIRQQLVVES 178
QY 177 GLDVIHNGDVIRIPFPVQLEMPDVNRLVDPDPNTHRSIGGEVYPTPPYNTGCLHL 236
Db 179 QIQSVRNGECDAEFVPLXDELFLAKSPCR-----ENLRY---FYHAGVSRY 221
QY 237 IHDCVIAPMAVALRVNRTAVARGAHLAFDENEGAVLPDITYTFYFQSSSSGTTARG 296
Db 222 LFEAHYTSLAQALRVVRVPGILALERQSF---HDYKLP-----KYECREPVATGHRG 273
QY 297 ARRDVNSTSPSPSGFERLASIMADTALLHAEVIFNTGIEETP-TDIKEMPMFLGM 355
Db 274 A-----GICSLLIYDSVATELAVSYGLSFLFEPQECALISYDKMPITEGC 319
QY 356 EGTLPRLNALGYSYTAARVAVIGAVVESPNALYLTEVEDSGMTEKDGPGPSFNRFYQ 415
Db 320 ETEPQREVALTFQNAKQAVHVSOLFSGNSVLYLAPQKQASN--PQSGGNVYNSPFG 377
QY 416 AGPRLAANPOTDRG-----HYLSSQSTGSSNTEFSVDYALICGAPALILRLLYLER 470
Db 378 HGLACLAEPQKENGLESPFGVPASALSSN---YSLHILAVYASFQMLARHCYLOF 434
QY 471 CDGAFITGHDALIKYVTFDSEIPCSICEKHTRPVCAHTYHRLRQMPRG-FGQATRO 529
Db 435 AQHQKSSNNGSVNVPYVIGVGAANTPMCELCRGSPPACVNTLFLRLDRPPVAVASVRKD 494
QY 530 PIGVFGIMNSYSDCDPLGNVAYLILKQPGDQTEAKAIMQDYATLIERLIDLEQER 589
Db 495 PYVATGAGA-YNDLDIAGNFANY---RDKDEESNQEEREKEFTYMQVOTQVL-----ER 545
QY 590 LLDRGAPCSSEGLSSVIVDPPTFRILDTLARIIEQTTQPMKVLVYTRDYKIRHGISFA 649

```

[illegible]



QY 965 TSTHLYRAAMNORFWVLGISTISKYGAAGNNRVPQACN---WSGLNGCKNVCPLFT-HD 1020  
 Db 969 -----GRPIVYPLVANNKKTGTINGNTQITQCGNLGTFKRGVDRNLIPDSTGR 957  
 QY 1021 RTRETIACPRGCTCP-----VTGSSSNRET-----TLSDVVRGJIVSGAMVLAIVA 1071  
 Db 958 KÖNNSSYRRHRHVFMTVPVAHLVKKNSNINLNFVEETIRKKNVONLFEDKD---NLINHD 1014  
 QY 1072 TV---VAVACARAGIMAFECWELSTLDEFFLAFLELHELQIITLLETPTWTEALEAV 1126  
 Db 1015 NVLELVKGLTQSGENTEDDQIÖFYLGEEYIMSDETWSRFQITTDGAPWSEVNTKVL 1073

## RESULT 14

103105

major single-stranded DNA binding protein - alphaherpesvirus 1

C:Species: alphaherpesvirus 1

C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 24-Nov-1999

C:Accession: T03105

C:Author: A. J. Pflanz, R. Fleckenstein, B.

C:Title: Primary structure of the alphaherpesvirus 1 genome.

A:Reference number 214840, MIMD 97404659, PMID 9261371

A:Accession: T03105

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1127 &lt;ENS&gt;

A:Cross-references: EMBL AF005370; NID:92337967; FIDN:AA05639.1; FID 92337973

C:Superfamily: herpesvirus DNA binding protein

C:Keywords: DNA binding

Query Match 9.5% Score 596; DB 2; Length 1127;

Best Local Similarity 22 R% Fred No. 1 R=36;

Matches 281; Conservative 209; Mismatches 542; Indels 202; Gaps 43;

QY 11 PTCPIGTYACRVEDIDLEISFLAARSDSLALFLMRNITVEKTPSSIAVNSCA... 68  
 Db 20 PTCGCFIYIPETLHNKHLSDCKYAEAGFSLPLHGVTEVAVPNNKAYKKKID 79  
 QY 69 RTTLAGAGITLKLTSHEFVSVEFHGKQIVLPSSAANPLFRACNAAREFESKQGP 128  
 Db 80 MIT-----VSVKIST--FYNNKALIFHNVRKRESIFSGGLGSLCKEACDLPQY--VPT 129  
 QY 129 PVDKAVETTCAGETICRIGLIPENTLLYLVTALFEKPAFMCVPLHYGLDIVHNGIV 188  
 Db 130 PL-GEGSTDVSDICPPVMOE-KDAIMAVVITGEGERELHGLKLYLKSQMSHVMINKTEV 187  
 QY 189 TRIPLPVQILFMPDVNRLVPPFNTHHRSIGCGFYVPTPFYNTGLCHLIDCVIAPMAVA 248  
 Db 188 YKIPLYDEHDLF-----TKSSLRK-----LYLPVASEVLYLTLYTSLAQS 227  
 QY 249 LRVFNVTAVRCAAHILATDENHEGAVLPEDITYFYFUSSSSITTAKARRNDVNSIKF 308  
 Db 228 LRVNMAASIVAEIQGFHDKYKMAKIVSKFYELATVACQITL----- 272  
 QY 309 SFGSGFERRIASIMADTALAAEVIFNGITVRETPP-DIKEMPMFLGMGCTPLRINALGS 367  
 Db 273 -----MVIDVAEAELGLSYLSLSEFEAPQEKTVQDDYSDIFASCETDSDRLALSK 324  
 QY 368 YTAHVAGVIGAVFSPNSALVLEWDSQMTAKDQGPDS-----FNRYOGAFPHL 420  
 Db 325 WNAIQAIHILHQLPSTNSIYYNNVAKQ-----APITNSKVEVNNVNSIYLQHCILAN 376  
 QY 421 AANPDTRDGHVLSQSTSS--NTFESVYIALIGFGAPLLARLLFYLEPRTDAGAPFG 478  
 Db 377 LCEHTLFEIHSAPFAGAPASSIDGSSFTLOHLYAAAFSPNLLARWCYLOPQOQHKSTIL 436  
 QY 479 GNG-IALYVYGTITSELICSLCKHIRVCAHIVHRLKQMKPKGQALIKQIPGPGM 537  
 Db 437 NFAYVITEVGSAAVSPV-CSLDSGQPCVQINTLFLYRLKDRPPVLYGSRDPPVITGI 495  
 QY 538 NSQVSDCPGLCNVAYVYLIILKKPGDQTFAAKATVMDYATLTKRLFLIDPQERKILDGCAAC 597

Db 496 TVVFNELDFLGNFASHPRKDDQDQNTQ---ETPRATYQJNLQTLLEKLEAGLVQ---SPV 550  
 QY 598 SDEGL-SSYIVDHPFFTRILLDLFAFTEQTLTPMKVLVETPDYKIFBELSEATHSMALT 656  
 Db 551 ADEGAGSGSMNLEKVFPPESDIDSLVDAEAKFINTMK-NNVNPKESTIKGVSHVIOYN 609  
 QY 657 PDYSGAPCPITNFIIVKPRHLAVVODLALSQCHVYGGQVQGVQVRNRPQPOP-VIARRP 714  
 Db 610 CNTWAPCSLMNLYKRSILTIIDIALP-----ISTVSEENPAGQKPNEMILKLN 663  
 QY 715 VDLFNGSFTSTPSTIVTTLSEGPVSAFPPLTGDPAFAGPTDQD-----LA 759  
 Db 664 QTLW-----TNPKSPFDIKGVITGTEMKVYNAKCSDFPDVDAALNNMSPYKGVQVILA 717  
 QY 760 PYSVEYIPDIPYKKNRVFEGNNTNLSEAPARLVGLASVYQGEKVD--MAGALGFL 817  
 Db 718 KAVVVALKNIKKRNKLFLFSG--TSKSEHY-----NAFLKANKRDVILLAGPYVKEL 768  
 QY 818 KPHGLIFPGMPPNKSPPNQMFWT-LIQRNMPADKLTHERITTLAAVKKP----- 869  
 Db 769 NSFHRLF-----PNKLT-SCLYMSNFKKKOIPCVPSAE-----ALNKEFSYINN 817  
 QY 870 TEEYALINFLPPTCIGELAPFYMANLILKYCHSUYLINTLSTITGAKRPDPSVL 929  
 Db 818 SKQFEVNMIDVVPDVSIVYAKQKLNNAIILKAGCQIUFYAVTISHIPKVOE---TCAL 873  
 QY 930 HMRKQVTSADQIETAKLLEKTEMLPELMTTAFSTILVYAAVN---GPPVAVYLSI 986  
 Db 874 ETPHYVLTSSVD-----SVEDYNNVQNLKAL---TYNSSRESANILARSRLVTLPLVAV 926  
 QY 987 SKYHGAAGNNRVQACN---WSGLNGCKNVCPLFTFDRTRRFTIACPRGCFICPVTC--- 1040  
 Db 927 NNYTGIAAGLFLGSANLIYFM-FRIVKMLLDLCLVKKLNLYMKKYLIMFELVSNLL 986  
 QY 1041 -PSSGNRET-----LSDQVPGIYSG-NAMQALYATVAVRAVRAQAHMAF 1086  
 Db 987 KPSYHOGIAFELEIVKKTQSLLEDDAQDEVLN-----RVACLVKYSIAGACADITL 1039  
 QY 1087 DD-WLSLTDDEFARLDELHDDIOTLETPMTEGALAEVAKLDEKTTAGDDETPNLA 1145  
 Db 1040 DITQFLVSYSMSSENILEKLDQ-LPELVAPKTHMAESVTK-----SGTVELELVQF 1091  
 QY 1146 FNPDSCEPHDITSNVLINISGNSISGTVPGIKR 1179  
 Db 1092 VAFE-----EEGVKLTSMDSKSVVGGKKR 1116

## RESULT 15

T44001

major DNA binding protein [imported] - human herpesvirus 6

C:Species: human herpesvirus 6

C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000

C:Accession: T44001

C:Author: Y. Mukai, T. Nakano, K. Kagawa, M. Chen, J. Mori, Y. Sunagawa, T. K.

J. Vitol, 73, 8053-8063, 1999

A:Title: Comparison of the complete DNA sequences of human herpesvirus 6 variants A a

A:Reference number: 227332; MIMD:99412319; PMID:10482554

A:Accession: T44001

A:Molecule type: DNA

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Cross-references: EMBL AB021506, NID:94995977, FIDN:BA076262.1, FID:94996029

A:Experimental source: strain HST; pop. variant B

R:Domínguez, G.; Damhaugh, T. R.; Stamey, F. P.; Dewhurst, S.; Inoue, N.; Pellet, P. E.

Vitol 73, 8040-8052, 1999

A:Title: Human herpesvirus 6B genome sequence: coding content and comparison with hum

A:Reference number: 227334, MIMD:99412318, PMID:10482553

A:Accession: T44188

A:Molecule type: DNA

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Cross-references: EMBL AF157706; FIDN:AA06339.1

A:Experimental source: strain 229; variant B

Cloned from  
Accession: 041  
Clustal Family: herpesvirus DNA binding protein

Query Match: 7.8% Score 490.5; DB 2; Length 1132;

Best Local Similarity: 22.8% Field No: 26 28;

Matches: 278; Conserved: 180; Mismatches: 535; Indels: 273; Gaps: 54;

```

01 2 ENKQIVETVGTGAYVAVRVEDIDIEELSPAASTSDALDIPMRLVEKPTSS 61
02 3 DNEIVVSAIVSAIVAVYVPRKEKELDVLSVLSLMKRNPPVLSILMLVENDPSTT 62
03 6.2 LAVVSCAPPTGLAGAVITLKLTTSTPPSVYFPGKINVL-PSSAANPLTRAGNAERF 130
04 6.3 V-----KPTLFNFGDILIKLITSPVCPFGHGTROIQMAEDHDLRLCEQTRQK 115
05 1.21 CSKTCGPGVDAVETGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
06 116 HUSIEVPAPKAYIDKA--LCSAVGRDSDVIGIVAGCGKELF-----ATLIL 165
07 181 -----VHINHGIVIRITFPVQLEMPDVNMLVDPENITHRSIGEGVYPTPYNTG 232
08 166 PVEHIOVOVGCEYGVKILYLSATLPETR-----ETLSQCTHETQKCHFLPA 216
09 2.33 EYHLTHKCVATPMAVAPRYNAVAVAGCAILLATEENHEAVLPFGTTTYFQSSSGTT 292
10 217 TSEETLPPYVFTSMGTLTRESNKEILDAGLKQPTQDQDQVTKLAPKRYL----- 266
11 293 TARGARRNVNSTSKSPSGCFERKLASIMADPTALAEVFN-----TGIVEETPTD 346
12 267 GLSGQKISAVEKD-----FLMLVSSV TELSFHVAEYLDVYD--PSQI 309
13 647 ---KKIMMELQMGCTLPRIALGSYTARVAGVIGAMVSPNSALVTEVE-----D 394
14 310 MNPENMHLIRNSETHAKMAKILNKLHLSHIAVLIFAPNSILLYCSKLAFIPNVKQAFN 369
15 395 SINTFAKKDQPGPSEKRFQVAGPULLAANDQDRDRI--VELSQSTGSSNTEFSVDYALI 453
16 370 SVMTQ-----ELDKSLSPFNALSSITRDVYNDNRKILKQDSTGKDKFSANHLAYA 422
17 454 TCGCATLALRLPYRDKACAPFGCHALXVYCTPSEI-----PQSLCEKH 503
18 423 CATCPQLTSYVWNLRKMSV-----YNAGNATETIYNHLVNSANLCFECDDK 470
19 504 TRIVVAHTTVHRIKQMPREFGQATROPICVPGIMNSQYSLQDILGNVATYLLIRKPGDQT 563
20 471 TQVSTIGTAMAVVGIKIPAIKPKVKKPIVMSKSPYAVAVDILGSPG-----RKPVSRL 525
21 564 EAAKATIMQTYATLEKFLDLEQERLDRGA-----PQSEGLSVIVDHPTR--- 613
22 526 KETGKQKQNTLS-----LDKGFVSOIFPYCKKNSLIDVYTGHDTPNVS 570
23 614 KILDEHAKIHQITTPKAVLEVIRDKYI-KKGLSEAFISMALIPDYSCAPCI 667
24 571 KRIEVSITIMDTGCTHEVVS--KCIENMRFGUTPREQIENCLQSPFNVDITPVATAFSP 627
25 668 TNPVAKRTHLAVDILASGCHGVFGQGVENRPNQFQPIARRKPVILF-----N 719
26 628 LLEFATYKVLILVIONAL-----VASGHVADKPTGNSISKMLVOQYOSLYGTFHSSYLK 683
27 720 GSEFTSTSTITVLTSEGVSAAPNPTLQD--DA---PAGR---TFDGLARVSEVIRDIR 770
28 684 KQFLNTRIVK-----ASNVDMQIIDGCHLYKSGKYVKTITQAKICRLSMQCLDRFR 735
29 771 VKNRVVFSQNTINLSAARLVGLASAVQOKRKYMDLHCAICPILIKQFHGLIPRQMP 830
30 736 IKNRPNKSSKTAIINPYEKKNV-----KHKKNPLSGCISFLFLFYHDKLF----- 781
31 841 ENSK SPINQVFWITLQORQMTADKLTHEETIT--AAVKKRTEFEVAAINPILNPPCIGR 888
32 782 INVKISLCLHMQRPILANNVPRKTLIDGNPREVKTFIKFASITNTYDEIDILIDQPECLST 841
33 889 LAQFYMANLILKYCHHSYLL--INTLS--ITG-----ARRPRDS--SVLHMIRK 934

```

```

DB 842 FIDCYFENKRLSALGFPHDYILISLGLSKLVTQNPVLFVVLVDKQPKFSSIGUELYVVK 901
QY 935 DVTSADILETOAKALLKKTENLPLPLMTTAPTSHLYVAAMNORPMVVICISISKVHCAAG 994
DB 902 LVLDG--VPNPVYASLSKEBN---FQITFS-----RSLVTFGLLEKFSVLA- 944
QY 995 NNRVQAG--NMSGLNK--GRNVCP-----FTFDPTFRFTNPPRGGFTQPTGSPSSN 1045
DB 945 NREYFOGOLWIGGSGVDKMLNPTSSAIDPFPRMR--QKTLIAKPSVIV-----KKVR 998
DB 1046 RETLSQY--VKGIIYS-----GGANVLAIYATVAVAGAKAQMADFDMLSLTD-E 1096
DB 999 RETIMFDEYVKGKVSIVENLTNDIDPELLIIAEVNR---DREDKPTMDMLFEVDGRE 1055
QY 1097 FIARD-----LEHLDQIIGTLETPWYEGALAV-----KILDKKTIAGDC--- 1138
DB 1056 ALAASIMLKLNHLVDMNVRDPSIA-NIQSVFEAVSSNDAPVDFSEILAEEDQASGLK 1114
QY 1139 --ETPT 1142
DB 1115 CDELET 1120

```

Search completed: March 28, 2003, 13.39.02  
Job time: 60 secs





QY 592 DRGA-----PCSEGLSSVYDHPFRRIIDJHARIBOTTPQMKVLYETRDYKIRBGI 646  
 DB 597 DEKALAHAGTGSAS--TGAVDAQSFNIIISTIKITTEGACAEQFMTELEVEDEFKIREGL 654  
 QY 647 SEATHSMALTEPDYSGACFPITNFIKRTHLAVVDLALSOCHCEVYGVQVEGRERNOF 706  
 DB 655 ADANHTMSISLDPSYSSSCPVTSFISKRITIFAVLDLVLISQCHCLFYGGSVGRNRNOF 714  
 QY 707 QPVLRRRVDLFGNGEISTRTSTVTLSECPVSAAPNPLTGQDAPAGTPEGDILARVSEVI 766  
 DB 715 QPVLRRRFLDLNGSGFITAKTVTVVSDSGVAPNLITPSSPEPTKDYGDGMARVSMEL 774  
 QY 767 RD1YKKNVVEVSGNCTNISEAARALVGLASAVQROKRVKVDLHGAIGPLIKOPGLLP 826  
 DB 775 RDLRIKNVLEFNGSNGANSEARAVAGMASVRRPEKSNILNCAVGLVQFIKVLFP 834  
 QY 827 RGMPPNSPPOWFTVTLTKRMADKLTHEITTLAAVKRTEDEYVAINFINIPPTCI 886  
 DB 835 RGHPRGIDTPNQQWFTLQRMOPARLISKEDIFTALIKRFSHYSAINFINTPNMI 894  
 QY 887 GELAQFYMANLILKXCDHSOYLINTLSTITGARPRDPSSVLTWHIRKQVTSAAQIETQ 946  
 DB 895 GELAQFYANLVKXCDHSOYFINTLITAVGSRPRPAVLANIKRTINCAQSDVEPPA 954  
 QY 947 KALLKTEINPLMTTASTSHLVKAAANNORPVVLGISISKYHGAAGNNRVQAGNNSG 1006  
 DB 955 QEVILQULSSNAAMVGTFASTINMRYVMDORPMVYIGLISISKYSGAGNNRVQAGNNSG 1014  
 QY 1007 LNGKKNVPLEFTRETRFIACPRGFGICPVYTGSSGNRETTISDOYRGITVSGAMVQ 1066  
 DB 1015 LNGKKNVPLEFTRETRFIACPRGFGICPVYTGSSGNRETTISDOYRGITVSGAMVQ 1074  
 QY 1067 LAIYAVYRAVGAQAQIHAFDMLSLTDEFLARDLEEDHDIQITLEPTVEGALFAV 1126  
 DB 1075 TAVESVYVLTALGARTQHLAVDEMLIGLVDEFLLASLDALNAAVVOVF--GEMSVGAQOMI 1133  
 QY 1127 KILDEKT-----TAGDGTPTNIAFNPDSCHSHPTJNSVLNISSNSISGTVPLKRP 1181  
 DB 1134 RTMCAQTNMGVSTGDS-----AFDFGALVGDANUSSTTEMGPA--SSSAPAGQKRPH 1185  
 QY 1182 EDELFDLSGIPIRKGNITMEM 1203  
 DB 1186 PDDILFDGAPPEKKSGILPDM 1207

RESULT 3  
 Q89549 PRELIMINARY; PRT: 1203 AA.  
 Q89549  
 Q89549: 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DE 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 GN 0129.  
 OS hOvine herpesvirus 1.  
 OC viruses, dsDNA viruses, no RNA stage; herpesviridae;  
 UC Alphaherpesvirinae; Varicelloviruses.  
 QX NCBI\_TaxID=10320;  
 RN [1]  
 RP SOURCE: FROM N.A.  
 RC STRAIN-COOPER;  
 RA Schwyzer M., Vleck C., Lowery D.E., Bellio L.J., Meyer G., Mista V.,  
 RL Submitted (MUG-1996) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SOURCE: FROM N.A.  
 RC STRAIN-COOPER;  
 RA Meyer G., Vleck C., Lowery D.E., Pastorek P., Thiry P., Schwyzer M.,  
 RL Submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SOURCE: FROM N.A.  
 RC STRAIN-JURA;  
 RA Schwyzer M., Vleck C., Lowery D.E., Bellio L.J., Meyer G., Mista V.,

RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JURA;  
 RA Schwyzer M.;  
 DR Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: 278205; CAB01596.1;  
 DR EMBL: X94677; CA64336.1;  
 DR EMBL: AJ004801; CA606104.1;  
 DR InterPro: IPR000635; Viral\_DNA\_bind.  
 DR Pfam: PF00747; Viral\_DNA\_dp.1.  
 SQ SEQUENCE 1203 AA; 127409 MW; 8299D64966a9654f CRC64;

Query Match 53.6%, Score 371.5, UB 12, Length 1203;  
 Best Local Similarity 53.6%; Pred. No. 5, 8e-263;  
 Matches 652; Conservative 197; Mismatches 340; Indels 27; Gaps 16;

QY 1 MENTOKTVVPTGPIGYYVACREDDLEISPLAKRSTDSIALPLPMRNLTVEKTPTS 60  
 DB 1 MDAAKTVALAGCPACGFYVVCDAARLDLKLALAAKRSQSLAVLJLVKGLTETAFAP 60  
 VY 61 SLAVSGARTLLAAGITLKLITSHFYPVSVEFHGKHVLEPSSAAPNLTRACNAAREF 120  
 DB 61 NVAVVAGTKTGTGLGAGLAKLPSHYHNVFVFGCEKRLASTAPNLTRACQARRR 120  
 QY 121 GFSRCQGPVDCAVETTGACFICRLGLEPENTILYLVVALKREAVKCNVPLHYGGLD 180  
 DB 121 GFSARAGAPVDAVETTADICTRAGVASEPLVLTATTEAKREYVMKNTLHYGGAAP 180  
 QY 181 VHIHQDVIRIPLPVOLPMPDVNRLVPPFNTNHRISICEGFVYPTPEYNTGICHLIND 240  
 DB 181 VQVAGFAVRVPLVYQJLYMPVNVNLPFNPAKQCAIGQLAYPRPFYNAALCCHLCY 240  
 QY 241 VTAPMAVNLRVNNAVANGAHLAFDENHGAVALPPTITTYFPOSSSGTTARGAPPN 300  
 DB 241 VGPAAVALRVKTVTLVAAGAHLAFDESHGAVALPDPVCTVF--DQPRGRCGRCSSAG 299  
 QY 301 DVNSTKSPSGCFERRLASIMADTALAEVIFNTGIEEPTDIEKPMFICMGCTLP 360  
 DB 300 ADPGAAKAAAPGVERRLASVMAADTAVSIEAMSTSVDEDEVACVDDPMLGGAADAA 359  
 QY 361 RLNALGSTYARAGVIGAMVESPNSALYLTVEYDSGMTEAKGCGPSPNRRYOFAGPHL 420  
 DB 360 KIDALCAVYGRLAGLVGAVVFSNVLHMT EYDDCGAADAADKGA--AAGCHRPYQIAAPYA 418  
 QY 421 AANPOTDRDGHVLSQSGTG-----SSNTEFSVDYALICGGAFLARLLTYLERCDG 474  
 DB 419 AGNPRCKDKRKLPTGAGPAVSIINGACGEFALDHALACGCPQLLAHMLTYLERCDG 478  
 QY 475 AFTGCHG--DAIKYVITGTFDSEITPGLCEKHTRPVCAHTTVHRLRORMPRGQATROPICV 533  
 DB 479 AFAGRNDDALKYVASTLEGDVPGLGSRDNRHAAHTLHLRPLHRLPRFGAPTPSPGLV 538  
 QY 534 FGTNMSOYSDCDPLGNVAPYULILRKQDOTFAAKAMIMODTYATLEIRLITDIOEHLDR 593  
 DB 539 FGTNMSAISDCDVLGNVASYSLAPRG--ADEAKRSIMQTYRAAENVLALEIDQOHLAH 597  
 QY 594 GAPCSSEGLSSVYDHPFRRIIDTLRAITEQTTTOFMKVLYETRDYKIREGLSEATHSM 653  
 DB 598 DAOSAAQ--LEAITDHAISFRGLAALIMTVBOATEFAFVGLVEDDFKREALVIANHTL 656  
 QY 654 ALTEPYSAGCPITNFIKRTHLAVVDLALSOCHCEVYGVQVEGRERNOFVLRARR 713  
 DB 657 SLADDPYTAACPATASFRRSVLAHVVDLALSOCHGIFGCGVGRERNOFVLRARR 716  
 QY 714 FVDLFNGCFITSTRITVTLSECPVSAAPNPLTGQDAPAGTPEGDILARVSEVIKRIYKN 773  
 DB 717 FMDLNGFLITRTIVTLAEAAVAPNLAAGVLRPAKUMIGUISKYSLEFVKRMRKYN 776  
 QY 774 RVVSGNCTINSEARALVGLASAVQROK--RVDMLHGLGFLDKQHGILFPGGMPN 832  
 DB 777 RVVPSAGNANSEARAVLGLAGAYOKPESGCVNLTLSGPIGLVKKOPHRKILFPNGKPRG 836



069101  
ID 069101 PRELIMINARY: PRT: 1197 AA  
AC 069101:  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, last annotation update)  
DE DNA binding protein ICP8.  
OS Herpes simplex virus (type 2).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Simplexvirus.  
OX NCH1\_Taxid=10310;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KN:  
RX MEDLIN=93228441; PubMed=8385914;  
KA Tsch Y., Tanaka S., Liu Y., Mori K.;  
\*Nucleotide sequence of the major DNA binding protein of herpes  
RT Simplex virus type 2 and comparison with the type 1 counterpart.\*  
BL Arch. Virol. 129:183-196(1993)  
EMBL: D10658; BAA01507.1;  
Interpro: IPR000635; Viral\_DNA\_bind.  
DR Pfam: PF00747; Viral\_DNA\_dp.1.  
SO SOURCE: 1197 AA; 128470 MW; AA3ADA75B8865BFE CRC64;

Query Match 47.9%; Score 3012; DB 12; Length 1197.  
Best Local Similarity 49.0%; Pfam No. 6.6e-214;  
Matches 595; Conservative 212; Mismatches 368; Indels 40; Gaps 16;

QY 4 TQKTVTVTTCIGYVY--AGRVHDLDEETSPILAASTSDIALPLMNTIVKKTFTSS 61  
DB 7 TTTVKKVPGPGVGYVGRACPAEGLEL--LSLSARSGADAVAVALLGLTVESGEAN 64  
QY 62 LAVSGARTTGACAGCTTCTTSHYTSVFVHGCKHVLSSAANLIRACMAANERG 121  
DB 65 VAAVVGSGCTGCGTAVSILKIMPSHSVSVYVFGCRHLAPSTQAINLRLCKRARRHC 124  
QY 122 FSRGCGPVDCAVETTGAEICTRLDEPENTILYVLTALFEKAVFMCNVELHYGSLIV 181  
DB 125 FSDVAPRCDLKHETTGDAICERLGLDPRALLIVITGCFREAVGINSITPLHICMDKV 184  
QY 182 HINHCDAVRIPLFPVQLEMDVGNRIVPDPFNTHRSIGCFYVPPPYNTGCHLIDCV 241  
DB 185 TIGDAEVRHPIYPLQMTMPDESRYADPCNHSIGENENTPLDEFNRLARLLEAV 244  
QY 242 IAPMAVLRVNRVTVAVARCAHILAFDENEGAVLPDIIYTYFOSSSGCTTARGARRND 301  
DB 245 VGIAAVVALKARNVDAAKAAHLAFDENEGALLADITFAPEASQC--KIQKQK-- 299  
QY 302 VNSTKSPSSGGERLSTMAADTALIAEVFNFTGTYEETPTDIKEMPMFGMECTLP 361  
DB 300 --DAGNKGACGFEDHILASVMAGDIALLESIVSMAVFPDEPPDITTWILKQDETPAK 357  
QY 362 TNAISYTAKEVGVIGAMVFPNSALYDFEVDSCMTFAKDCGPGSPFRFYQAFPHLA 421  
DB 368 AGAVGATYLARAGLVGAMVFTNSALHILTEVDAGAPDKDKSK--PSTPFLVPGTHVA 416  
QY 422 ANIQTRDRCHVL-----SSQSTSSNTSPSVLYLALICGQAPILANLELYLPCDAG 474  
DB 417 ANIQDRDRCHVVGVEGHPILAPVQCTQ--EFAGEHLAMLCGSPALLAKMLFLPHCDG 475  
QY 475 AFTGRHGS--DALEVVNTFTSETFTSLAEKTKPVCAHTIVHLRQMPFEGVATKPLV 533  
DB 476 VLVGQEMVNFVYVAVNSCOTIVNCNCTFTTHACAHHTILMRIRAHHPKFAASAARCAIGV 535  
QY 534 FCTMNSQSTACHPGCAVAYVLIIRKQDQFAAKATMDIYVYKATLHPIR--LIDFQFLDIP 593  
DB 536 FCTMNSAIVLQVLLVLAFAFSAKLR--ADGSEMTRTIMQETYSAAETRYVALELQVVO 594  
QY 594 GAPSSEGLSSVIVHPIVHPIVHPIVHPIVHPIVHPIVHPIVHPIVHPIVHPIVHPIVH 653  
DB 595 AVPTALGRIETIIGKREALHIVNNIKQIVDRVHPIVHPIVHPIVHPIVHPIVHPIVHPIV 654  
QY 654 ALTEPFYSACPFITNLVKRTHLAVVVDLALISQCHICVGGVVEGRNRPNGQVPLRR 713

1197 AA; 128470 MW; AA3ADA75B8865BFE CRC64;

655 SLSEPTTTPVPLDLLARSKSLAVYGLLALSYHSAVTAQSVESGKRNPNFQPLRR 714  
QY 714 FVDIFNGCFISTRSTVTLSG--PVSAIPNPVIGQDAVACHPIDGILAVSVEVIRDLVK 772  
DB 715 VMDLFNNGELSAKTLTVALSEGNAICAPSLTAGOTAPAESSEFGDVARVLTGPELRYK 774  
QY 773 NPVPSGNTNLSEARALVGLASVYQDEKPVIMHNA--SLKLVGFHLLFPPKMPN 832  
DB 775 SRYLFGASANASEAKAKAVASIOSAVQKPKRVDLILGLPGLIKQFAVIVPCKP 834  
QY 833 SKSPNQMWTLLQKNMPEADKLTIEITIAVKKRETEVAALINFLPTCIGELAP 892  
DB 835 SNPNQMWTLAGRNQRLPARLSKEDIETIAIKKESLDYGAINFLNPNVSELAMY 894  
QY 893 YMANLLKQDHSOYLINTLSITTCARRPDPSSVLMWIKRVTSADILEQAKALIEK 952  
DB 895 YMANULRYVTHSTYFINLITAVIAGSPPSVJAAAMAFQ--SVGLEAGAFALMDS 951  
QY 953 TENLPETMTAFTSTHUVAAAMQREPMVYGISISYHAAAGNNRVEAGNMGKGN 1012  
DB 952 LDHHPGAMTSMFASCNLRYMAARVYVIGLISITSTYTGAGDQVFOGNNMSILGKN 1011  
QY 1013 VCPLEFDETRPPIACPPRGFTPYTSPSSINPFTLSTVVPGLIVSYAMVLAIVAT 1072  
DB 1072 TKSLSGPRVQVLEDMALLLEDEYLSSEMEFTTALKEGHEGEMSTDALAEVHDEAL 1131  
QY 1129 LDEKTAGDGETPNTAFNEDSCPSHDITSNLTNSGNSGSTVPGKRPPEDELF 1188  
DB 1132 VSLDZAGH-----VNNQVIGQDTHHAASFGIAAA--AAAGAVARHPARHDDPFC 1183  
QY 1189 LSGIPKIGNTMEM 1203  
DB 1184 ESTPEPK--TITLDM 1196

RESULT 6  
ID 09P252 PRELIMINARY: PRT: 1190 AA.  
AC 09P252:  
DT 01-MAY-2000 (TREMblrel. 13, Created)  
DT 01-MAY-2000 (TREMblrel. 13, last sequence update)  
DT 01-JUN-2001 (TREMblrel. 17, last annotation update)  
DE Homolog of HSV-1 glycoprotein.  
OS Turkey herpesvirus.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Marek's disease-like viruses.  
OX NCH1\_Taxid=10390;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HPRS24;  
KA Kato K., Jiang H., Izumiya Y., Cai J., Tsushima Y., Miyazawa T.,  
Kai C., Mikami T.;  
\*Identification and Transcriptional Analysis of the Marek's Disease  
RT Virus SeroType 2 Genes Homologous to the Glycoprotein B (UL27), the  
RT ICP18.5 (UL28) and the Major DNA-binding Protein (UL29) Genes of  
KT Herpes Simplex Virus Type 1.\*  
BL J. Vet. Med. Sci. 0:0-0(1999).  
EMBL: AB024711; BAA83753.1;  
Interpro: IPR000635; Viral\_DNA\_bind.  
DR Pfam: PF00747; Viral\_DNA\_dp.1.  
SO SEQUENCE: 1190 AA; 130339 MW; D3291E3EAAFF5AD CRC64;

Query Match 44.4%; Score 2796.5; DB 12; Length 1190;  
Best Local Similarity 45.3%; Pfam No. 1.7e-216;  
Matches 546; Conservative 245; Mismatches 29; Indels 29; Gaps 13;

QY 6 KTVVTPGIVGYVACRVEDLEELISPLAARSTDSDLALPLMRNLTVKKTSSLA 65

```

10b 5 KSVKTTGGPITGVYAAFTSSMPAEDELSEFAKSNCEDEALLPLVSGTLEADPMNVAAV 64
07 SCARTTGLAGAGITLTKLTSHFYPVSVFPHGCKHVLPSAAANLIRACAAARHRRPSAC 125
10b 65 AGRTITGLAGAGITLTKLTSHFYPVSVFPHGCKHVLPSAAANLIRACAAARHRRPSAC 123
07 126 QCPVPGAVETTGAEICTGLALEPENTILYLVVTALEKAEVEMCNVELHGLDVIHNIH 185
10b 124 YSPVATSEETTGQICNIGNNPQETMLYLVTLELFRKAVYLCNSYLHYGCGVSSING 183
07 186 GIVIRLPLFPVGLFMDVNRKLVDPDPNTHHRSIGQGVPTTPHYNGICHLIRKCYIAIM 245
10b 184 VIVRRPLPLFPLVDFDRKVASDPESTKPRALGKALPKAEYNDSDRLHGLYLSIA 243
07 246 AVAIRKRVNTAAVAKGAHILAFDENHGAVALPDPITVYTGSSSSGCTTAKGARRNVNST 305
10b 244 AVGIRKRVNDALAKCAHILSPDENHGLTLPADTAFNATPIESASKSYKAKGCKGMEL 303
07 306 SKPSPSGEPHPLASTMAALTLIAEVLINTGITYEETPDLEKMPFETGMEGTLPRLNL 365
10b 304 S GCGEERTASTIMASDALTILENVLATISVYESIHDVKKMPLVCSPTGCTDRFVAL 359
07 366 GSYIAPVAVGIAHVSINGALYCTEVHSCMIFAKD GCTPDSINRREYQAPDPIAAM 424
10b 360 SAVMAHVNATVCAAMPSSNSVITTEVREBASSAKDKTSTTAASYRFRFOIAPHLASNP 419
07 425 GTPHGHVLS...SGTSSNNEPESVDYALITPCGADILARIIPYIRPGACATTCCH 480
10b 420 LIDRCKKPLHOLSKAISASHEVSLDYLIIACGCPDILARPLPYLRKOGSOGACH 479
07 481 G DALKYATGTDSLEPSTLEKTRPVCAHTTVIRIRQMRPRFGATGTPRGVFTNMS 539
10b 480 LIDRCKKPLHOLSKAISASHEVSLDYLIIACGCPDILARPLPYLRKOGSOGACH 539
07 540 QYSDKPLKAPVATVLEKPKQTPAAKATMDQVYKATLEKLEPIHJQKRLIDKAGACSS 599
10b 540 NCGVNALGSAVPESTLKR...SEGETSRSMQDPTKKTVEPMAKLEKSGITLTCEPDM 597
07 600 RCLSSVIVDHPFKRLIDTLAKIKOTTTPQPKVIVETIKYIRIGLSATISMALTRIP 659
10b 598 TTAVALIKRCKSPKMAISIMANIIPSEACQIMKNIIEIRVNIHROIGATHTLSLIEP 657
07 660 YSGAETPTNEFKRIHLAVVODLALSQCEVFGQVEGRNRFNOFVLPBRREVDLEN 719
10b 658 YSSGICVATSLSRRTILAVGDMAISQSMIMHGQVYARHPRPQGVIPRRVINDION 717
07 720 GGFISTRSITVILSECPVSAVNPITIGUAPAGRTFDGDLARVSVVIRDIKVKHRYVESC 779
10b 718 AGFTTSKNITVTELDQJISVDPKSKQIHDPLAVIHESGLVYTFEELREFEKVKMYEVG 777
07 780 NPTNLSAANRAHVGIANAVGOKRKHVIMLHAGLGLIKOPHGLIPRGMIPNSKSPRP 838
10b 778 CVSSIVSATKSKRLAKMIFAYVYAKAMHVINPGLGFAIKRHTDLPVAKMNGTINA 837
07 849 QWNTLILQHNOMPAUKLTHEELITTAIVAKRTEBYAANFINLPTCIGELAOFYMANLI 898
10b 848 LFWHTLILQHNOMPAUKLTHEELITTAIVAKRTEBYAANFINLPTCIGELAOFYMANLI 897
07 849 LKYLTHSGLVILNLSITIGARRKDSVILHMKDVIISADEIIOKALLKELINLPE 958
10b 848 LKYLTHSGLVILNLSITIGARRKDSVILHMKDVIISADEIIOKALLKELINLPE 957
07 959 LMTAFTSTHIVRAAMNODPVAVLIGISISKYHGAAGNNVPOAGMNSGICNCKNCPILT 1018
10b 958 AMCAAFSSNINAVGPMASKPVPVAVIAVSIKHYGMAOSTKVPQSCMNGNIMGRNCSIMS 1017
07 1019 FIKTRRTIAPRGFTIPVTPGSSNRETTILSDVPCITVSSNEMVQATATATVAVRAV 1078
10b 1018 FNRHRRVAVACDVRGVPVSHQGTGSSGILKETTIVBARAAVLSREGCAPHAAVYMAIALKMG 1077
07 1079 AKACHMADIMNLSITDDEPLAARHETTHQILITLFTWTVGCALEAAKILDEKTTAGIC 1138
10b 1078 DWRHMEIDMMEITINAVISSLIDELNKO...VEGEGGMSVDAAAILAKEVDMKSLPIL 1136

```

```

07 1139 EPTNLAENDESCPESHDTTSNVLTSSGNSISSTVPGKRPPEDELDLSIPIKHK- 1197
10b 1137 DQPT---PNTDIDENRERAD-----GPT---LPEITIKKPSD---VPLERVPRKAP 1183
07 1198 NITEM 1203
10b 1184 GLSDM 1189

RESULT 7
09QTB9 PRELIMINARY: PRT: 1190 AA.
AC 09QTB9;
DT 01-MAY-2000 (TREMblrel, 13, Created)
DT 01-MAY-2000 (TREMblrel, 13, last sequence update)
DT 01-MAR-2002 (TREMblrel, 20, last annotation update)
DE UL29 protein.
GN ORF 37 OR UL29.
OS Marek's disease virus serotype 2 MDV2, and
OS Gallid herpesvirus 3.
OC Viruses; dsDNA viruses, no RNA stage, Herpesviridae;
OC Alphaherpesvirinae, Marek's disease-like viruses.
OX NCBI_TaxID=95353, 35250;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES-Gallid herpesvirus 3 (serotype 2); STRAIN-HPRS24;
RA Jang H., Cai J., Izumiya Y., Murakami Y., Mochizuki M., Song C.,
RA Lee Y., Kai C., Takahashi K., Mikami T.;
RT "The complete DNA sequence and transcription map of the unique long
RT genome region of Marek's disease virus type 2."
PL Submitted (Feb-1999) to the EMBL/Genbank/DBS databases.
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES-Gallid herpesvirus 3; STRAIN-HPRS24;
RA Izumiya Y., Jang H., Ono M., Mikami T.;
RL Submitted (Oct-2000) to the EMBL/Genbank/DBS databases.
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES-Gallid herpesvirus 3; STRAIN-HPRS24;
RA Izumiya Y., Jang H., Ono M., Mikami T.;
RT "A complete genomic DNA sequence of Marek's Disease Virus Type 2,
RT Strain HPRS24."
RL Curr. Top. Microbiol. Immunol. 0-0-0(2000).
DR EMBL: AB024414; BAAB2925.1; -
DR FMBL: AB049735; BAB16539.1; -
DR InterPro: IPR006335; Viral_DNA_bind.
DR Pfam: PF00747; Viral_DNA_bp_1.
SU SEQUENCE 1190 AA; 130357 MW; 05589C36CA0B0E9 CRC64;

Query Match 44.4%; Score 2793.57; DH 12; length 1190;
best local similarity 45.2%, Fred. No. 36-216;
Matches 545; Conservative 236; Mismatches 396; Indels 29; Gaps 13;

```



```

QY 306 SKPSPSGEGERFLASTIMADTLHAIEVINTGIEETPTDIDKEMPMFIGMEGTLPRNLAL 365
DB 304 S-----GGTERRKRLASTIMADATLITINVAITSVEESIPDVKMPLTCSPTITDVEAL 359
QY 366 GSYTARAVAGVIGAMVSPSALYLFEVEDSGMTEAKD-GGPPSPFNKYFQAPPHLAAP 424
DB 360 SAYMARVAGLVGAMVSSSVIYMTVEGASSADGDTSTTASFRFQIAPPHLASNP 419
QY 425 QTPRHGHVLS-----SOSTGSSNTEFSVDTALICGCGAPILARLLPYLRKQACATQCH 480
DB 420 LTRKGRPIRGEDLSKATISASPEVSLDYLACGSPULLARFLPYLERKDEGSAACHH 479
QY 481 G-DALKYVGTGDSSEIPECSICEKHTPRVCAHTVHRLROMPREGQATROPICVFGMNS 539
DB 480 DLDYKFFVSAMDADVPCHDCKASKRIYCAHTTKRLERHLRPFQYOMKAMGLFQSMIN 539
QY 540 QYSLCDPLGNVAPYLLLRKPDQTEAAKATMODIYRATLERLIDLEUPERLDRAPOSS 544
DB 540 NYCDVVALGSYAPFTLKR--SEGFTSRVMDTYKLTVERVAKALEKGLTCEBDPADM 597
QY 600 KGLSSVTVHPHPRRLIDTLRAKHOTTFQPMVLVETRYKIRKGLSEATSHMALTDP 659
DB 598 TPADATIRGKSEFMRLSTMANITSEAGQMLMNLTEIREYNIKEGLUGAFTLSAIEP 657
QY 660 YSAGAEPTINFLVKRTHLAVVLDLALSCVYGVGVGEGRNFRNFPVLRFRFYLEN 719
DB 658 YSSGICPVLSPLSKRTIIVAVQUMALISQCSMIMHQOVERNRNRTQFOALRRRVIDLON 717
QY 720 GGFSTRKSTVTLSEGVVSAPELTQUDAPAGKTFEGDLARVSEVYRIDIKVKNVSES 779
DB 718 AGFTISKNTVLEDDQIISVPDPSQHDPLAVHMEGDLVKTFFELREKFKVKNKVFVS 777
QY 780 NCIN-LSFAARALVGLASAYQKQKRVMLHGLGFLKQFGLLPPRRMPNSNSPNP 838
DB 778 GVSSTVSDATKSLAGMIEAYQPPAKAMHVLNPLGFALKRHYTLQLEPVYKMGNGTTPNA 837
QY 839 GWFETLLQNPADKLTHEEITTAIVKRFTEEYAINFINLPICIGELAGFYMANLI 898
DB 838 LMFVILLQJNOILPACILISDAEHTSFIRKPTNSYADMTINISPTCFGLAOFYIANT 897
QY 899 LAYTTHSOTVLTITSTTGAPRPPSSVLMIRKQDVSADITQAKALEKTENLPE 958
DB 898 LKYSCKHKEFINISALVAISKPRPVAWLPEHILITGSDVAAVQULLENICDHE 957
QY 959 LMTATSTHILRAANONKRVAVLGISIKYHAGCNKRFVAGNMSGCNVCLPI 1018
DB 958 AMCAAFSSNINLGFIMASKPFVYIAVISIKYHMASTKVEVSNNKNTIMGSRNVSIMS 1017
QY 1019 FDRTRFTIACPGAFTCPVYTGSSGNPETTESDQVRGIIYVSGAMVQIATVATVPVAV 1078
DB 1018 FNRTHRFVACPRVGHVSQDTGFSGCKETTLVDRARAVISECGAPHAHVYMAIAKMG 1077
QY 1079 ARQUMHAFDMLSLTUETLAKLELHDQIITLETPTVVEGALAEVAKLDEKTTAGDG 1138
DB 1078 DRYROMELDDMMITIDAVYSSLIDELNKQ-VEGCGGWSVDAALILAKMVMMSKPLPL 1136
QY 1139 EPTVNLAFNDUSCEPSHOUTSNVNLISGMSISITVPGLKRPEDEDLFPLSGIPKHG- 1197
DB 1137 DQGT---FNYDALDENTERRAD-----GFSI---LEPTLKRPLSD--VFOLFVPEKRAP 1183
QY 1198 NITMEM 1203
DB 1184 GLSVDM 1189

```

```

RESULT 8
Q966P0 PRELIMINARY: PRT: 1191 AA.
AC Q966P0:
DT 01-MAR-2001 (TREMHLrel, 16, Created)
DT 01-MAR-2001 (TREMHLrel, 16, Last sequence update)
DT 01-JUN-2001 (TREMHLrel, 17, Last annotation update)

```

```

DE U129 single stranded DNA binding protein-like protein.
GN MDV042
OS Turkey herpesvirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Marek's disease-like viruses.
OC NCBI_TaxID=103390;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD5;
RX MEDLINE=20392152; PubMed=10933706;
RA Tulman E.R., Alfonso C.L., Lu Z., Zsak L., Rock D.L., Kutish G.F.;
PT "The genome of a very virulent Marek's disease virus."
J. Virol. 74:7980-7988(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MD5;
RA Tulman E.R., Alfonso C.L., Lu Z., Zsak L., Rock D.L., Kutish G.F.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases
DR EMBL: AE243438; AMG1422.1;
DR InterPro: IPR000635; Viral_DNA_bind.
DR Pfam: PF00747; Viral_DNA_bp.1;
SQ SHQUNCE 1191 AA; 130714 MW; 9AD425780F807711 CXC64;

Query Match 44.3%; Score 2788; DB 12; Length 1191;
Best Local Similarity 45.1%; Pred. No. 8.4e-216;
Matches 542; Conservative 237; Mismatches 395; Indels 28; Gaps 12;

QY 1 MENTOKTVYPTPTGLGVVACREDDLDLEISFLAANSTISDIALPILMKNLYEKFTS 60
DB 1 MDGVGVSVKLGCGPPIGIVYATPKQSVAPDEBALIIMAKSNCCDVAFLVAGLTVESDFW 60

QY 61 SLAVVSGARTTGLAGAGITKLTSHFYPSEVFEHGKHVLPSSAANLRACNAAREF 120
DB 61 NVAAVACTKTITGLSGGTTIKLVPYTHHPCVFPYGGDCIKPCKAINTKACILAEHF 120

QY 121 GYS-AYSSPAFTFEETGGQICELEMDAONVMYLIVTELEKVIYLCNSLFHGSDV 179
DB 121 VHNHGVIRKLPFVQVLMHVDVKNLVPDPNTHRSIGCFVYPTTFYNTGCHLHIC 240

QY 181 VTNINNAVPRIPYPLHLVLPDEFPIHNEPSEKPRALGGAIPKAFNDSTCRLLGY 239
DB 180 VTNINNAVPRIPYPLHLVLPDEFPIHNEPSEKPRALGGAIPKAFNDSTCRLLGY 239

QY 241 VAPMAVAVRVNVTAVARAANHLAFDENHEGVALPDDITYTYGSSSSSTTARFAPN 309
DB 240 VLSITAGLKVKNLIDALAKAAHLCHDNNHSTLLPADITFIATTAAR---TKKQSKM 296

QY 301 DVNSTKPSPSAEPFRLASIMAACTLAHAEVINTGIEETPTDIDKEMPMFIGMEGTLPR 360
DB 297 GKREGSDVS-GGGERFTASTIMADATLAINVISASVDEDPIDPVKMPYICNPVGYAD 355

QY 361 RLNAAGSYTARVAGV-GAMVSPSALYLFEVEDSGMTEAKD-GGPPSPFNKYFQAPPH 419
DB 356 KIELSLATMARVAGLVGAMVSSSVIYMTVEGAGSTEGEISTVAPSTFRFQIAPPH 415

QY 420 LAANPQTRDGHVLS-----SOSTGSSNTEFSVDTALICGCGAPILARLLPYLERCDAGA 475
DB 416 LSNAPLVDRKGRKPVSSNLSKTSASQSEVSLDYLILACGCPQLARLLPYLERKQCGA 475

QY 476 FTGSHG-DALKYVGTGDSSEIPEVSL/EKHTPRVCAHTVHRLROMPREGQATROPIC 534
DB 476 KACHHDLDYKFFVSSAIDADMPCELCDKTSIKYCAHTTKRLVYRLPKFGYOMKAMGLF 535

QY 535 GTMNSQVSTPDPICGNVAPYLLLRKPDQTEAAKATMODIYRATLERLIDLEUPERLDRA 594
DB 536 GAMTNKCDVVALGSYAPFTLKR--SEGFTSRVMDTYKLTVERVAKALEKGLTCEBD 593

QY 595 APCSSSEGLSSVVDHPTFRRLIDTLRAKHOTTFQPMVLVETRYKIRKGLSEATSHMA 654
DB 594 DPTVMAADANIRDGNSEFIRAISTMKNITSEAGQMLMNLTEIREYNIKEGLUGAFTLS 653

QY 655 LTRPPYSGARVPTNPLVKRTHLAVVLDLALSCVYGVGVGEGRNFRNFPVLRFRFYLEN 714

```

654 LAVERASDQIPIALISREK11AAVDDMALSCSIVMGGQVEARNR1QFOAVLKRHV 713  
 715 VDFENGEFTSPSTLTAT SPSPSADNPITLADAPAFRTTEGLARVSEVIDIRKVR 774  
 714 LEONNGFTSKITITVLEQOUCVDPSPKSDVSVISNNEGDLVKVTEIEFELKVRK 773  
 775 VESNNGCTN1SEAAARAVTASAVQROKRVKMLHQAFLKOPHQLFPRCMKPS 833  
 774 VLEGGIADVAASEATSKSLAGAEAVQRPKTMHVLNGLPFAVKRHITLLEFDVKMPG 833  
 834 KSNPQWMTLLQORNMFKLTHEITTTAAVKRTTEEAALNFINLPRTGIGLAEFY 893  
 834 ATTNALMFLKRLKQIPALISKEEDMSLTKPKFKSAIDMNYINI SPICGDLAOFY 893  
 894 MNALILKYDHSVYLINTSLITLTKRRPDPSSVLHMKDYTSADIEOKALLEKT 953  
 894 LANTILKYSKHKIEPIINTISALVASKRPDPAIVLPWIERPTTKQGDVAPAOQLISM 953  
 954 ENLPDELMTAFTSTHLYKAAANORPMVVLGISTSKTHGAGANNRVQACWMSGINGCKNV 1013  
 954 SDHRDIAWTFSSSTNI VSSIMTKRPFVIGISTSKYHGMAGSTKRVQSGMW:NMGGPNV 1013  
 1014 GLETFDTRKRETIARPRGECIPVTGSPSSGNPFTLLSDQPGITVSG:AMVOLAIYATV 1073  
 1014 GSIASDRIHRVYMTCPKQCFVABOP1PSSG1KETITLIDKRVNVLSPSSAPPAAVTMIA 1073  
 1074 VAVAVARQAQMAFDLMLSTIDDEFARLDEELHDOIIQTLETPWTEGAEVAKIIDEKT 1133  
 1074 LKVVAVRPMQFLEPMFTTNEEYISLLHELNKQ-VEEAEKGMNDADLAKEMVMA 1132  
 1134 LACDCEPINLAFNHSCESHDSTSNVNLNCSNLSGIVPQIKRPHDELFDLSGIP 1193  
 1134 MSIPITKPT-FOVACLENELHMAQOTISETN-----LKRPNMNVFDELP 1179  
 1194 IK 1195  
 1180 EK 1181

RESULT 9  
 VQDPOR  
 10 VQDPOR PRELIMINARY: PRT: 1190 AA.  
 01 MAR 2001 (TREMURC. 16, created)  
 01 MAR 2001 (TREMURC. 16, last sequence update)  
 01 DEC 2001 (TREMURC. 19, last annotation update)  
 06 H29 single stranded DNA binding protein.  
 06 H2937.  
 06 Meleagrid herpesvirus 1 (herpesvirus of turkeys).  
 06 Viruses: dsDNA viruses, no RNA stage: Herpesviridae.  
 06 Alphan herpesvirine: Marek's disease-like viruses.  
 06 NCBI TaxID: 37108.  
 06  
 06 SEQUENCE FROM N.A.  
 06 STRAIN: F0126.  
 06 MEDLINE: 20578442; PubMed: 1134410.  
 06 Alonso C.L., Tulman E.R., Lu Z., Zsak L., Rock D.L., Kutilsh G.K.,  
 06 EMML: AF291866; AAG4767.1.  
 06 InterPro: IPR000615; Viral DNA\_bind.  
 06 Protein: P400747; Viral DNA\_bp: 1.  
 06 SEQUENCE: 1190 AA; 140054 MW; 4450CAF5PF00928 CRC64;

Query Match: 44.2%; Score 2781.5; DB 12; Length 1190;  
 Best Local Similarity: 44.5%; Pred. No. 215;  
 Matches: 548; Conservative: 245; Mismatches: 391; Indels: 35; Gaps: 14;

QY 1 MENTOKTVEPTGIGVYACRVEDLDELSPFLASTDSIDALIPMNI:VEKFTS 60  
 1 MDVGKRVKLVGPISGVMSKDAVVDDELSTFAACISSESLIDLALITVESFEVR 60  
 QY 61 SLAVSAGRTLAGAGITLKTTHSHFVPSVFHGGKRVLPSSAENLRACNAAREP 120  
 61 NVAVAAGIKITGIVNGSTIKETVIRHHPVAVYGGACINPCNAUNILKACVAREP 120  
 QY 121 GFSKQGPDPDGAVEETGAEITRLGLEPENTILLYLVVALAFKAVKCVNLFHYGGLDI 180  
 121 GYCEPSAP---GSVEISGKQICBGLDAOKTMYLVVTELFKATILCSNFFYGGSDS 177  
 QY 181 VTHNGDVILPLPPVOLFPMVDVNLVDPFNTNHSIGCFVYPTFVYTGCHLIDHC 240  
 178 VTINDAEVRIRPLYLPLVLVLDENRVSNSEFSEPRALRGAVWPEAFDLCRLLYGY 237  
 QY 241 VIAPMAVLEVRNVTAVARAAMHAFDENEGAVLPDITTYTF--QSSSGITTARAG 298  
 238 VLGTDVAGLVARNIDAVAKAAHLCPDENHGHILLPADTTFATPTADYDGGQKQK 297  
 QY 299 RNDVNSTKSPSGGFERRLASTMAADTALHAVENTGTGYEETPDIKEMPMIGMEGT 358  
 298 RDGAEE-----TGGGVERKRTASLMSADATLAIENVIVSSYEEDIPYKMPITFCNVGY 352  
 QY 359 LPRINALGSTANAVAVIGAVPSPNSALYLVEVNSGMTFAKDG-POHSPNHYQFAG 417  
 353 TDVEALSAVAGVAGLVGAVMFSNSVVIYMEVGEASSDGGKESVPAPSEYHFQIA 412  
 QY 418 PHLAANDQTRDGHVLSOS- TGSNTEFVDYALITGFAPLALLYLEGDA 473  
 413 PHLAANDLVRODKPI SGDALPKIASTSSVSLDYLILACGCPQOLAFPLFLEGDC 472  
 QY 474 GAETGSG DALRYVTCTPDSLPSCLECKETPRVAAHTVHLRPMPEFPAATGEC 522  
 473 GANACHHDLDTVAKLVSADADMPCELDKTSRYCAHTTKVLEKVLPRFGYQMRKMG 542  
 QY 533 VPGTMSOYSPDPPDNYAYLILKRPQDTEAAKATMOTRYATILRLPILDKERLLD 592  
 533 LFGAMTNNGDVALLSYAGFTLKR--SEGEASRSMODITRLRYERVAKALEKELLA 590  
 QY 593 RPAQSSSELSVVDHPPTPLIDTLFARIEGTTQFMVIVETPYKIPESLSEATHS 652  
 591 CDDPMNAIPADASI KQOPSTHAVSTMRNILEGASGLAMNILEIRKYNIREGICANHT 650  
 QY 653 MATIDPYSGAPPTNFTNFKRTHLVAVDIALSQHVYVQVSPNFPNFPVLP 712  
 651 TSLTIEPSSGFPVLSFLAKRTTIAVIGFMALSC:SMLEHGAQVEARNR1QFOAVLKR 710  
 QY 713 RPYDNLNCGFISTRTITVILSEGVASAPNPTLGDPAACKTIDGILARVAVVILHDIRK 772  
 711 RVELOHAGTTSKDTITATLEDOHVAVPDPSRSQYDPTVNMESDGLRVTETIRELKV 770  
 QY 773 NRVFSGNCT-NLESAARAVTASAVQROKRVKMLHQAFLKOPHQLFPRCMKP 831  
 771 NRVFSGNCT-NLESAARAVTASAVQROKRVKMLHQAFLKOPHQLFPRCMKP 830  
 QY 832 NKSPPQWMTLLQORNMFKLTHEITTTAAVKRTTEEAALNFINLPRTGIGLAEFY 891  
 831 NGAVPNAHMEWTLLOKRNQIPOTMISKEDEKALEIKFTVYVADMYNINISPTGFDLQ 890  
 QY 892 FYMANILILKYDHSVYLINTSLITLTKRRPDPSSVLHMKDYTSADIEOKALLEKT 951  
 891 FFLANILIKYSKHKIEPIINTISALVASKRPDPAIVLPWIERPTTKQGDVAPAOQLISM 950  
 QY 952 KTEDELMTAFTSTHLYKAAANORPMVVLGISTSKTHGAGANNRVQACWMSGINGCK 1011  
 951 TLADRKDAWCATFASATINIVGSVSTKPLIVIGVISTKYHOMAGSTKRVQSGNMGNI 1010  
 QY 1012 NVCPLETFDTRKRETIARPRGECIPVTGSPSSGNPFTLLSDQPGITVSG:AMVOLAIY 1071  
 1011 NVCSLSMDFRTHRYVACPCVGVVAEHTFSSGLKATLMDKAKALISBGAAPHAEEVY 1070  
 QY 1072 TVAVARQAQMAFDLMLSTIDDEFARLDEELHDOIIQTLETPWTEGAEVAKIIDEKT 1131



\*Marek disease virus encodes a basic leucine zipper gene resembling the fos/jun oncoproteins that is highly expressed in lymphoblastoid tumors.\*

RT: protein. Natl. Acad. Sci. U.S.A. 89:4042-4046(1992).  
 RT: [2].  
 RT: SEQUENCE FROM N.A.  
 RT: STRAIN: CAI  
 RT: MEDLINE: 20283955; PubMed: 10824954;  
 RT: Lee L.P., Wu P., Su D., Ren D., Kamil J., Kang H.J., Witter R.L.,  
 RT: "The complete unique long sequence and the complete genomic  
 RT: organization of the CA strain of Marek's disease virus."\*  
 RT: Proc. Natl. Acad. Sci. U.S.A. 97:6091-6096(2000).  
 RT: PMID: 10478065; AAF6764.1;  
 RT: InterPro: IPR000645; Viral\_DNA\_bind.  
 RT: Pfam: PF00747; Viral\_DNA\_bp.1;  
 RT: SEQUENCE: 1191 AA; 140946 MW; AC04A015B80797A3 CRC64;

Query Match 44.1%; Score 2710; DB 12; Length 1191;  
 Best Local Similarity 44.1%; Pred. No. 1.7e-209;  
 Matches 545; Conservative 246; Mismatches 400; Indels 42; Gaps 16;

1 MENTKIVTPTPLIYVACVVEDELEISF-----LAARSTSDALLPLMNTLV 54  
 1 MGVKSVVIGGPIA IYMR-----PNVAPPMNCGLAKSNCGDAVPLVAGLIV 53  
 55 EKFTSSIAVVGAKITGLACAGITIKITSHPVSVFVHGGKHAIVSSAANITRACN 114  
 54 ESDPVWNVAAVAKITIGSGSTITLKVPTHYHPCVFEGGDCIKPCKAPNLRKACD 113  
 115 AARERPGSKTGGPIVDAVETTCATCTRIQLEPENTHIVAVLAKFAVHNCVAFH 174  
 114 LAKRATGYS ASSAPPAFETTCOOLICALMDAONMAYLVVTLKRAVITLCSPH 172  
 175 YAGLIVHINIGDVRILEPVULFMPDVNRIVPDEFTHIRSTAEFVYPTPTVGT 234  
 174 ECHSVVITINNAVRRIPYPLHIVIPDENRITNEFSKRALGCAIMPAFVINDSLC 232  
 235 HLLHKVIAPPAVALKRVNIVAVAGCAHLADENHGVLPIDITTYTFOSSSCOTTA 294  
 234 RLIRVYVLTAVGVKRVKIDIAKAAVHDPDENIEHTLLEADITFAFAAE---TT 289  
 235 RLARRNDNSTSKPSGSEERRIASMAADALHAIVFNGIYEETPIDIKEMIPG 354  
 240 KQSNMKAKEKSDVS GGVYKRIASIMASIAITAEVINSVYHDPIDYNAKMTIYCN 348  
 455 MEGTIPRLNALGVTAFAVAVIGAVFSPNSAVLYLEVEDSMTAKD-GAPSPSENF 413  
 449 FGVYADRIEALSAVAKAVACIVGAVFSSNSVYIMFVGGASTGTETSTIAPSPR 408  
 414 GFADGHIAANVQIDKHGVLS----SQTGSSNTEFSVIALTGGPGAPILAKLIFYLE 469  
 409 GFADGHIAANVQIDKHGVLS----SQTGSSNTEFSVIALTGGPGAPILAKLIFYLE 468  
 470 KQVACAPTCCHD DALKVYVTFPDEIRFISLAEKTRVGCATTHRIKQHMIRKCOATR 528  
 469 KQVACAPTCCHD DALKVYVTFPDEIRFISLAEKTRVGCATTHRIKQHMIRKCOATR 528  
 529 OPTVFGTNSQSYKVDPLANAVPILIRKPGQTEAAKATMGTTPATLEPIFIDE 588  
 529 GAMLEGGAMTNNVGVNALVSAVSPSTIKR--SPGASHKSVMDQVRIIVRRMKALRKE 586  
 589 RLIDGAPVSSSGISSVIVDHPFRRLIDTPARITFOTJQMKVIVETROKLEHGL-S 647  
 587 GLTIDTPTNMAADANIEDNSFTALSTMNITIESEASQLMRNLTEFRENITPGWG 646  
 648 FAHSMALTFIDYSGAPPTINFAKRTHLAVVQIALSQCCHVYGOVFGCRNPNQ 707  
 647 DANHLSIAVEYASGICVIALISKRITLAVVQIALSQCSTVMGOQOVFAHNFPTQ 706  
 708 FVLRKRVLDLNGGFTSTRTVTLSRIVVSNPNTLQVDAAGKTFQGLAKVAVEVIR 767  
 707 AVLRKRVLDLNGGFTSTRTVTLSRIVVSNPNTLQVDAAGKTFQGLAKVAVEVIR 766

QY 768 DLKVKKKVVFSGNCIN-LSKAAKAVIGLASAVQROKRVDMHGAIGLLKQFHLLFP 826  
 DB 767 ELKVKKKVLEFGGIIACASASATSKRIAGVYEAQPTKTHVYVNGDFGVAKRYHLLFP 826  
 QY 827 RGMPSKSPNPQWFTLLQPNQMPADKLTHEETITAAVPEETEEVAAINFINLPPT 886  
 DB 827 DVKMPGATPNAIMFWLILLRNOLPACILSKFEEDSLFKKFTKSYADNVAINISPT 886  
 QY 887 GELAGYVNNILIKYVTHSHLYLNTLTSTITVAPRPPSPSVJHMLPKOVTSAADETQA 946  
 DB 887 GDLAQYVNNILIKYVTHSHLYLNTLTSTITVAPRPPSPSVJHMLPKOVTSAADETQA 946  
 QY 947 KALIEKTENIPLEMTAFSTHIVRAAMNORPVAVIGISIKHYGCAAGNNRVQACNWSG 1006  
 DB 947 QQIASMDSHKDQWCATFSSNTNVSIMTKPFVYIGISIKHYGCAAGNNRVQACNWSG 1006  
 QY 1007 LNSGKVCPLFTFDKTRKELIACPRGGLCPVTGPSSGNREITLSQVRCIIVSGAMVO 1066  
 DB 1007 IMCGRNVCISMSHPRHRYVMTCPRVGFVAFQPIFSSGIKETILLDRVHVLSESSAAPH 1066  
 QY 1067 LATYATVAVAGRAQHMAFDDMLSTIDDEFLARDEELHDOIIOTLETPVTEGALAV 1126  
 DB 1067 AAVYMLALKAVGDRVQOMELEDMETINDEYISLIDELKQ-VEEAEAGWADAAMTTLA 1125  
 QY 1127 KLIDEKTTAGDCTPYNLAAPNDSCHPMDYTSNVLNISGNSIGSTVQIAKRPDEHL 1186  
 DB 1126 KEVYNNAMSIPTDGP---FDFACDENLEGHADGQISETN-----LKRPN-NNMV 1172  
 QY 1187 FDLGSIPIKHN 1199  
 DB 1173 FDLPIPKRAV 1185

RESULT 12  
 Q90H63  
 ID Q90H63 IPELIMINAKY PRT: 999 AA.  
 AC Q90H63  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE DNA binding protein.  
 OS Gallid herpesvirus 1.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Infectious laryngotracheitis-like viruses.  
 OX NCBI\_TaxID=10386;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SA-2;  
 RA Johnson M.A.;  
 RT "Sequence of the infectious laryngotracheitis virus (SA-2 strain)  
 RT unique long region UL28 to UL41."\*  
 RT Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF168792; AAD56201.1;  
 DR InterPro: IPR000635; Viral\_DNA\_bind.  
 DR Pfam: PF00747; Viral\_DNA\_bp.1;  
 SO SEQUENCE 999 AA; 108477 MW; 084FAVTHC8H4674H CRC64;

Query Match 23.0%; Score 1445; DB 12; Length 999;  
 Best Local Similarity 30.5%; Pred. No. 2.3e-107;  
 Matches 351; Conservative 177; Mismatches 376; Indels 246; Gaps 18;

QY 13 GPGVYVACRVDDIDELISLIAKSTDS-DIALPLMNTIYEKTPSSIAVVSAGRTT 71  
 DB 22 GPGVYVACRVDDIDELISLIAKSTDS-DIALPLMNTIYEKTPSSIAVVSAGRTT 71

QY 72 GLACAGITIKITSHPVSVFVHGGKHAIVSSAANITRACNAARERPGSK-OGPIV 130  
 DB 82 GVGCGNCAIILSPCHSTTVITFHGPGHCIPYTSPTFNLIKCHDAKRRPGSSIPNCGV 141

QY 131 DGAVETTGAEICTRLGEPENTLVVLTALKEAVFMKCNVFLHGGDILVHINIGDVR 190  
 DB 142 PNALETGHEICKSINMDPKIMLYIVAVRPGCAVAVCNVILHGSVYVITAVINSKHTV 201

```

QY 191 TPLFVQLFMPDVN-RLVDPDPNTHHRSIGRCVYVPTPEYNTGLCHLJHDCVJAPMAVAL 249
Db 202 IFIYVQWYMDIDIALRICRNPFDNSKINGECITPTPLFNALNRKLHGAILGIGGSL 261
QY 250 RVRNVTAVARCAHILAFDENEGAVLPDITYTYFQSSSGTTTARGARRDNVNSTKPS 309
Db 262 RTKDLFAVARCAAMLAIFDGSFEGCVLTSUKPTTEL 307
QY 310 PSGFERELATIMADTALHAIEVINTGIEETPRD-----IKEMPMFISMEGLPRIN 363
Db 308 QNVEVERAAAGSLAEALATRLSLVSCAPY---PBDNGCGLEPYEEMPLECCDKTOMERO 364
QY 364 ALGSLTAVAVAGVIGAMVSPNSALVTEVEDSGMTFAKDGPG---PSFNHFOFAGPH 419
Db 365 ALDRSALAGVIGAMVSPNSALVTEVEDSGMTFAKDGPG---PGEASSGLTRFYLTCGH 419
QY 420 LAAMPVJDRDH-----VLSSSTGSSNTEFSVYVLLALICGAGALLARLLELCRDA 474
Db 420 LIGCGVAVYSGRWEGTQPSLTPQNTFYDPEHLAAACGSPHLSNIFYLEKCSKO 479
QY 475 AFTGGHGDALKYVTGTFSSEIFCS-----LCEKTRPVCAHTTVHRLRQMRPGQAT 527
Db 480 LSTRN---MNLISNSGKACPTCAVEPCRCNDESNHCTRTYMOHLRTFTPTPTPT 536
QY 528 KQPIGVCTMNSQVSDCDPLGNVARYLLLRKPDQTEAKAKATMODTYRATLEERLIDLE 587
Db 537 RSPAVIGAVAGDYTDCCILGNFAVYSLKBRAGD-GEPAKAMNDYTRKIGRWVFOYLV 595
QY 588 ERIDRGAPCSSEGLSVTVHPTFRPLDILRAREEDTTOFMVLYETROYKIREGIS 647
Db 596 EGLINKDGETHNSL-----AEIKTYFEKIFONFVGECSFISALSGVARYHYKHEIS 651
QY 648 EATHSMALTEPYSGAFCTPNTFLVKRTHLAAYVODLSQCHVYGOOVEGRNFRNOFO 707
Db 652 SIATFG-SPEYSSAYCPVLSLCAQTRSLIFODLILSOJHGTDTQPEKMRSSAM 710
QY 708 PVLKRRKVDLPNGCFILSTRITVILSRGPVAPNPTLQODAPAGTFCODLARAVEVIR 767
Db 711 PALSARNGMIDKGFGLSKYEPFESTVPCVNAPTMRPTEOALAOYEYSLSRGVLEIK 770
QY 768 CIPVKNPVVPSNCTNLSEAAFAVLGLASAYOROEKRVDMHGLGFLKOFHGLLEFR 827
Db 771 HFVKYKRVVFGSSSTNS----- 788
QY 828 GMPNSKSPNQWMTLLQKRMADKLTHEITIAVKTREYAAINFINDPPTCIG 887
Db 789 ----- 788
QY 888 ELAOFYMANLILKYCDHSOYLINTLSITGARRPRDPSSVYLWIRKDVTSADILETOAK 947
Db 789 ----- 788
QY 948 ALLKTEKTELPMLTAAFTSTHLVRAAMNQPMVVLGISTSKYHGAAGNNRFEQAGNMSGL 1007
Db 789 ----- 788
QY 1008 NGKKNVCPLEFTRRRTFLIACPRGGFTCPVTGSSSKRETTLSUQVGLIVSGAM 1064
Db 827 IGNSGSQSVNSALINDPYRKALACKRVTGLSSQVAVSGQENTIAQVAVSLDAGAS- 885
QY 1065 VOLATVAVAVAGARQOHMAFDMLSLTDEFLARLDELEHJDOITQLETFMVEGAL 1124
Db 886 -PTSIYAVLAVLEGEKMDIAETMAITDCKYLVNLVELREDISGS-QNGMSVATQAS 943
QY 1125 AVKLDEKTT 1134
Db 944 MLSELECKDT 953

```

RESULT 13  
ID 040913  
AC 040913;

PRHLMINARY:

PRH: 1132 AA.

```

DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, last annotation update)
DE ORF 06.
OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC Viruses: dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=37296;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97296220; PubMed=9151804;
RA Neipel F., Albrecht J.C., Fleckenstein B.;
RT "The genome of human herpesvirus 8 cloned from Kaposi's sarcoma";
RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: U93872; AAB62603.1; -;
DR InterPro: IPR002114; HPr_Serp_site.
DR InterPro: IPR001993; Mitoch_carrier.
DR InterPro: IPR000635; Mitoch_carrier.
DR Pfam: PF00747; Viral_DNA_bp.1.
DR PROSITE: PS00215; MITOCH_CARIER. UNKNOWN_1.
DR PROSITE: PS00589; PTS_HPR_SER. UNKNOWN_1.
SQ SEQUENCE 1132 AA; 125394 MW; CF88E41451919CBF CRC64;

Query Match
Best Local Similarity 10.9%; Score 683.5; DA 12; Length 1132;
Matches 301; Conservative 196; Mismatches 500; Indels 265; Gaps 53;

QY 11 PGPRLGYVACVEDLDELIFSLARSDSLPLMLKNTLVEKTFSTSLAVSGART 70
Db 18 PTPGCGVLYAYLTHNFPVIGEASLNGVPEAKVFSPLHLGLTVSDPFLNVKAVH- 73
QY 71 TGLAAGITLKLKTHSFYVSVFVGKGHVLPSSAANPLTRCANAREFGFSGRCGPV 130
Db 74 KRIDATTSVAKLTSYH--REALVFINTHLFQPIFGKGLKCRESLRELFGSTVEQOH 131
QY 131 DCAVETTGAEICTRIGLEPENTILYLVATLKEAVPMCNVFLHYGGLDI-----VH 182
Db 132 KQTL--WSPEDACPOLCANE-IFMAVITYEGEKERL-----YGGKLVVPQSQTPIVH 180
QY 183 INHGDVIRLPFVQOLFMMPDVNRLVDPDPNTHHRSIGSGFYVPTPEYNTGLCHLJHDCV 242
Db 181 IGEHAPFKPIAYDELDFGFSRAQEL-----CRPYNDISRYLHDSITP 222
QY 243 APMAYALVRNVTAVARCAHILAFDENEGAVLPDITYTYF-QSSSGTTTARGAAPPND 301
Db 223 TGLAQLRLKVEDSVIYQ-ASERQF--VHDYKIFMLVYAKDFPQCAHSRITDQ----- 271
QY 302 VNSTSKPSPSGFERELASIMADTALHAIEVIFNIGI-YETPTDIKE-----WPMPTG 354
Db 272 -----STLWVID-SLVAELGMSGLSTESIEGPODSCVELNVDTPWFEN 313
QY 355 MEGTLPRNLALGSIYARAVAGVIGAMVSPNSALVTEVEDSGMTFAKDGPGPSFRFQ 414
Db 314 CETPRARLAEVNHMAEQALHIGAOFLAANSVLYLTRV--ALPKQNRGDAMNNSFYL 371
QY 415 FAGPHLANPOTDRG-----HVLSSOSTGSSNTEFSVYVLLALICGAGALLARLLE 469
Db 372 QHGLGYLSEATVKEKNGASAFKGVPSALDGS--YTLOHLAYVASSFPHLLARMKYU 428
QY 470 RCDACAPFGCHG-DALKYVTVGTGFSSEIFCS-----LCEKTRPVCAHTTVHRLRQMRPGQAT 527
Db 429 FLPHHKNTNSQSYNVVDYV-GTAAPSQKCDLQCGCPRAVCINTLLEYRKMDRPPVLSNVK 487
QY 528 KQPIGVCTMNSQVSDCDPLGNVARYLLLRKPDQTEAKAKATMODTYRATLEERLIDLE 587
Db 488 RDPVAVITGTAGT-VNDELILGNFAIFRERREGRGNFVEDAP--KTYWQQLCONI----- 538

```



FT	VAR1ANT	SEQUENCE	1103	AA:	123272	MM:	056359648C8C27C03	CRC64,
SQ			356	S -> G				
	Query Match		10.4%	Score 652.5	DB 12	Length 1103		
	Best Local Similarity		23.9%	Pred. No. 3	1e-43			
	Matches 292	Conservative 194	Mismatches 535	Indels 199	Gaps 47			
QY	1	MENTOKTVIVPPIGLCYVACREDELDLEISFPIAAKSJDSLDLALILPKRNLTVEKFTLS	60					
DB	1	METTVNCSAPILGPAGIYIYQVLVDPEPPEASLSTINFTSAACILPLVTGTLVEGRNF	60					
QY	61	SLAAVSGATTTLACAGITLKLTTSHFYSVYFPGKQKIVLDPSSAAPNLTFRACNAAREF	120					
DB	61	NKK----APCOKIMESLTVKPT--KYFHAIYVFNHSMKIPRIFMGKGLQJCEFAKRLMF	114					
QY	121	GFSGCGPVPDVAVENTGAICTRLGLEBENTILVYVLTALKEAVFMQNVFLHGGDI	180					
DB	115	GFSSPKPADASKQMKSS-----THLPQNEKICGVVAASESKRILICGHIALMSAPKPT	169					
QY	181	VAIHNGDIVIRIPLFYQVLFMPVYKRLVDPPEFTNHRISICEGVYPTPYVYNGICHLIDHC	240					
DB	170	VYIINKKCEKPEVLHDIRLSSD--QL-----PYDENVSIMES	208					
QY	241	VYAPAAVALRVNNTAVAKGAHILFDEHNECAVAPRDIYTYFQSSSGCTTAAACARRN	300					
DB	209	YVTNVAQAVIRIDVIGCLILEADERAVTQYKAKIASLKY-----	249					
QY	301	DVNSTSKSPSGGFEFRRLASTMAADTALHAELYENTGIYEETP-TDIKEMPEIGMEGL	359					
DB	250	----PAEMCGACNILEYAAIDAASVSLAASHGLAFLEAFQENSPIDONYQWPMF--SPDAT	403					
QY	360	P--RLNALGSYTARAVAGIYGAIVSPNALXYLTVEEDSGMTFAKGGGPPS--NRPYQFA	416					
DB	304	PEGRIKALQEMAKKAHHVHQVLSANSVLYITTKYGGA--GAVAGGKDDMFLSTRYMOH	462					
QY	417	G-PRILAAPQTDROCHVLISSOSTGSS--MTEPSVDYIALICGHAPILAKLLPYLEKCD	472					
DB	363	GLGHL--VWQTFDENNTYVFCGVDPSCNMNGSOYVNHILAYAASMSPHILAKICYLQMCQ	420					
QY	473	AG--AFTGGHGDALEYVVTGTEDEISIPSLTEKHTPRVAHTVNHLEPKQMPR--FVQATFJ	529					
DB	421	NORLASTPGCTTVAPYVYTTTANTSM--CDLGGSDILPAVCLHTLFFHLKRPQVLYATQCKD	479					
QY	530	PICGVCTMNSOYSDCDPLGNANPAILLKPRIGYQTHAAKAMOUITYKALIERLFLDLEQR	589					
DB	480	FYVYVG-TAGPRINDEILIGNEGFS--KEKDDQ-----ENGARYSWMNLNL--TER	527					
QY	590	LIDRAPCSPSEGLSIVIDHPTFRILITDTPARILPQTTQPMKVIWPR--DYKILREGISF	648					
DB	528	LEELIGVTASTFA--QLITDVOVSFLVEFKKIDAVYDEETIRINDGLVAKNNINRK--ESKKS	584					
QY	649	ATHSMALTFDEPSCAFQPTNPLVKRTHLAVVQDIALSQ-----KCFVGAQVDSKHF	702					
DB	585	IAHVLIQLOCNNTMPLPPPIHNILFRSPFTTLODMSPFLCYTHNENPILNVG--TVSTWIM	642					
QY	703	RNOGFVLRRRRVDFLFCNGFTSTKSLIYILSKGSPSALNPLIGQCAPRGKIFDGLIARVS	762					
DB	643	SNHQNOLMSNKSITMFDKGLITGSMRVRVHE--TSSMENIADLPRQPAAT--PLSAQ	699					
QY	763	VFVIRIDIVKKNRVVFSQCTNLSFAARALVGLASAYOROKRVDMILGALCFLIKQPHG	822					
DB	700	IFAPPAKIKIKRRIIPSNSS--TSLSLASFYKACAEK-----PIAGPRMHPILOHK	751					
QY	823	LLEFGKMPNSKSPKQWFM--TLLOORNMP-----ADKLTHEETITAAVKRFTEEVA	875					
DB	752	QLP-----TTKICAVYIMQTFMSSKKILPQAGSTNAKKS--EPLSYLMTSIAHDE---	801					
QY	876	INFINLPPICIGELAOPPMANILKYCHSOYIINTITST-----ITGAVP	921					
DB	802	VNVLDYFTILMAVAKQRLNNAILRLCGQTFYATITNFLOPITLSMPSLDYPIHVG---	858					
QY	922	PRDSSVYHMRKIVTYSADILETOAKALLEKTENIPMLMTAFSTHLYAKMANQRIWV	981					

```

DB 859 PIVV... QDLITVMJMKDQSAIVLOSCHRD-----PAVCKMRPIVT 899
QY 982 IATISISKYHGAUNNREVEQAUNMSGUNGKNGCPLFTDR-----TRR 1024
DB 900 IPMWVKYSGSGNNSIFUSGN MGYFMGRGV-----DRKLLDIPRMKHYNTSMRRR 952
QY 1025 FIIACPRGCEIIVIGGSSGNRETTISDOVR-----CHVSGAMVOIAIYAVVRAVGAR 1080
DB 954 YAFASHPTESTL LTPGKQIVRTYQLEKTRKI IIDI IASGNAPE NLVLELVKCLGHE 1009
QY 1081 AQHMAFDOMLSTDOEFLAKOLELIDJITOTIETPTVEGAL EAVKIDEXKTAGDG 1138
DB 1010 CATTISD- DVEFYAGFSPWVNDVMEVLETLKASGCAFSESSVOALAPTSSPGTS 1063
QY 1139 ETPTNIAFNDS TEPSHOTT 1158
DB 1064 ESPVEVTDNITSLPETSIVT 1084

```

Search completed: March 28, 2003, 14:48:02  
 Elapsed time: 118 secs



GenCore version 5.1.4.P5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2003, 13:31:04 ; Search time 30 seconds

(without alignments)  
1663.201 Million cell updates/sec

Title: US-09-769-699-2

Perfect score: 6294

Sequence: 1 MENTOKTVTPGTPLGYVA

Scoring table: BLASTUM62

Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database: SwissProt\_40\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6294	100.0	1204	1 DNB1_VZVD	P09246 varicella-z
2	3620.5	57.5	1209	1 DNB1_HSVB	P28932 equine herp
3	3054	48.5	1196	1 DNB1_HSV1K	P17470 herpes simp
4	3049	48.4	1196	1 DNB1_HSV1L	P04296 herpes simp
5	3047	48.4	1196	1 DNB1_HSV1P	P17469 herpes simp
6	3015.5	47.9	1196	1 DNB1_HSV2H	P89452 herpes simp
7	3013	47.9	1197	1 DNB1_HSV2	P36384 herpes simp
8	2933	46.6	1186	1 DNB1_HSVB2	P12633 bovine herp
9	1094.5	17.4	375	1 DNB1_HSVB1	P03444 equine herp
10	669	10.6	1128	1 DNB1_EBV	P03227 Epstein-Bar
11	604	9.6	1128	1 DNB1_HSVSA	P24910 herpesvirus
12	492.5	7.8	1132	1 DNB1_HSV6U	P52338 human herpe
13	490.5	7.8	1132	1 DNB1_HSV6Z	P52338 human herpe
14	483	7.7	1131	1 DNB1_HSV7J	P52339 human herpe
15	479.5	7.6	1160	1 DNB1_HSV7C	P13215 simian cyto
16	477	7.6	1135	1 DNB1_HSV7A	P17147 human cytom
17	473	7.5	1193	1 DNB1_HSV72	P09117 herpesvirus
18	412	6.5	1191	1 DNB1_HSV71	P30672 murine cyto
19	121	1.9	1066	1 DNB1_HSV73	P09113 homo sapien
20	115	1.8	833	1 DNB1_HSV74	P18401 gallus gall
21	114	1.8	833	1 DNB1_HSV75	P18401 gallus gall
22	114	1.8	833	1 DNB1_HSV76	P18401 gallus gall
23	113.5	1.8	833	1 DNB1_HSV77	P18401 gallus gall
24	113	1.8	833	1 DNB1_HSV78	P18401 gallus gall
25	110	1.7	833	1 DNB1_HSV79	P18401 gallus gall
26	109	1.7	833	1 DNB1_HSV80	P18401 gallus gall
27	108	1.7	833	1 DNB1_HSV81	P18401 gallus gall
28	108	1.7	833	1 DNB1_HSV82	P18401 gallus gall
29	107	1.7	833	1 DNB1_HSV83	P18401 gallus gall
30	106.5	1.7	833	1 DNB1_HSV84	P18401 gallus gall
31	106	1.7	833	1 DNB1_HSV85	P18401 gallus gall
32	106	1.7	833	1 DNB1_HSV86	P18401 gallus gall
33	106	1.7	833	1 DNB1_HSV87	P18401 gallus gall

## ALIGNMENTS

34	105.5	1.7	1016	1	UVRA_DEIRA	Q46577 delinococcus
35	105.5	1.7	2114	1	MYB_MOUSE	Q99Y06 mus musculu
36	105.5	1.7	4540	1	DYHC_PASTE	Q27171 parametium
37	102.5	1.6	1902	1	P2P_LAC1C	P15293 lactococcus
38	102	1.6	1038	1	V120_VZVD	P09277 varicella-z
39	102	1.6	1275	1	RFBG_MYXA	Q50864 myxococcus
40	102	1.6	4590	1	FATH_HUMAN	Q14517 homo sapien
41	101.5	1.6	1478	1	BUD3_ASHGO	Q9h161 ashyba goss
42	101.5	1.6	1502	1	PIP_LAC1C	P16271 lactococcus
43	101	1.6	1229	1	CV43_TIRYB	Q99280 trypanosoma
44	101	1.6	1348	1	VCR2_COTUA	P52583 coturnix co
45	101	1.6	3649	1	ACVS_MOCIA	P27743 nocardia la

RESULT 1  
ID DNB1\_VZVD  
AC P09246  
STANDARD: PRT: 1204 AA

DT 01-MAR-1989 (rel. 10, Created)  
DT 01-MAR-1989 (rel. 10, Last sequence update)  
DT 16-OCT-2001 (rel. 40, Last annotation update)  
DE Major DNA-binding protein.

GN 29.

OS Varicella-zoster virus (strain Dumas) (VZV).

OC Alphaherpesvirinae, no RNA stage; Herpesviridae;

OC Alphaherpesvirinae; Varicellovirus.

OX NCBI\_TaxID=10338;

RP MEDLINE=86306657; PubMed=3018124;

RA Davidson A.J., Scott J.E.;

RT "The complete DNA sequence of varicella-zoster virus.";

J. Gen. Virol. 67:1759-1816 (1986).

CC -1- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA

REPLICATION.

CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).

CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN

FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

enterprises requires a license agreement (see <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CV EMBL: X04370; CAA27912.1; -

DR PIR: C27214; DMBE29.

DR InterPro: IPR00635; Viral\_DNA\_bind.

DR Pfam: PF00747; Viral\_DNA\_bp; 1.

KW DNA-binding; DNA replication; Zinc-finger; Nuclear protein.

FT ZN-FING 497 510

SEQUENCE 1204 AA, 132139 MW, DZFEFb5Dc0C674 CRC64;

Query Match 100.0%; Score 6294; PA 1; Length 1204.

Best local similarity 100.0%; Pred. No. 0;

Matches 1203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MENTOKTVTPGTPLGYVACRVERLIDLFISFLAARSTDSNALLPLMRNITVKKPTFS	60
DB	1	MENTOKTVTPGTPLGYVACRVERLIDLFISFLAARSTDSNALLPLMRNITVKKPTFS	60
QY	61	SLAVVSGARTTGLAGAGITLKTTSHPFSVFPFGCKHVPSSAARPLTRACNAAAREEF	120
DB	61	SLAVVSGARTTGLAGAGITLKTTSHPFSVFPFGCKHVPSSAARPLTRACNAAAREEF	120
QY	121	GFSCGCPVVDCAVEVTAETICRLGLEPNTILYLVATLFEKAVFCNVPFLHYGGLDI	180
DB	121	GFSCGCPVVDCAVEVTAETICRLGLEPNTILYLVATLFEKAVFCNVPFLHYGGLDI	180



```

QY 419 HLAAPQTDGRGHLV-----SSOSTGSSNTEFVSVDYALALICGAPLARLLFLERCD 472
DB 418 YVAGQPDQDKGRVLPHPADDOAAPIINSNOEFLSDYIALACGFCPOILARLLFLERCD 477
QY 473 AGATGT-GHGALKVVTSTFSELPSCLEKHTPRVCAMHTVYHRLQDMPFGQATROP1 531
DB 478 AGTEGGRNETALRYLANTLESDDVPCGCLNOATBPACAHITLHRLRORLPFGAGVRAPI 537
QY 532 GVFQGMNSQYSCDPIGNVAPYLLLRKPGDQTEAKATMODYRATLRPLFDLROHRL 591
DB 538 GIEGTMSAYSDCDVLGYNASTYALKRPND-NEAKSITMODYRATMELELQAKLI 596
QY 592 DR-----GAPCSSEGLSSVVDHPFRRLDILRARIEDOTTQFKVLETRDYKIREGL 646
DB 597 DKETLAQSPCSAP--TSVHDAQSFIGLNSIKDTICGAAGQFRLTLEADPFIRESL 654
QY 647 SEATHSMALTDPSYSGACPIITNFKVKTHLAVODLALSOCHCVFYGGQVGRNFRNOF 706
DB 655 ADANHTMSISLDPSYSSSCPTSTFLARTVFAVLDDLVLSCHCLEFYGSVSEGRNRNOF 714
QY 707 QPVLRRRFRVDLPNGCFISTRTSTVTLSSCPVSAHPPTLGDAPGRTFGDLARVSVEY1 766
DB 715 QPVLRRRRLDMLNGCFITAKTYIVTVSDSGVLAAPDLTPRASEPTIKVDGMARSMEL 774
QY 767 PDIRYKNPVFVSGNCTNLSEAAARLVLGLASAYROEKRVDMHGALGELLKQFGLLEP 826
DB 775 RDLHYKNRVLRKNGCANNSFAAPAVAGMASAYBRPDGKSNILLNGAVFLVQYHCVLEP 834
QY 827 KGMFENKSPEDPOMFTLLOKUMPAOKLTHEITTAIVKRFTEEVANININLPICI 886
DB 835 KGPFGIDTPPOMFTLLOKUMPAOKLTHEITTAIVKRFTEEVANININLPICI 894
QY 887 GELAOFYMANLLKYCDHSOVLINNTSTITGARRPRDPSSVLMIRKIDVSAADLETA 946
DB 895 GELAOFYMANLLKYCDHSOVLINNTSTITGARRPRDPSSVLMIRKIDVSAADLETA 954
QY 947 KALLEKTENPELWTATSTHLYRAANORPYYVLGISIKYHGAAGNNRYFOAGNNSS 1006
DB 955 QEVLRQSLNSNAAMTGTSTINMYVYVMDORPVYVIGLISIKYHGAAGNNRYFOAGNNSS 1014
QY 1007 LGGKKNVCPLEFTRTRFIACPRGFCICPVTPGSSGNRETTLSDOYKGIIVSGAMV 1066
DB 1015 LGGKKNVCPLEFTRTRFIACPRGFCICPVTPGSSGNRETTLSDOYKGIIVSGAMV 1074
QY 1067 LAIYAVYAVCARQAHAFDMWLSLTDDEPLARDLELDQIIQYLETPTWTCALBAV 1126
DB 1075 TAVFAAVYVLAHAGARTQHLAVLWLGVLDEFLAASLALNATVYDQF-GEMSVFAAUELY 1133
QY 1127 KILDEKTAG--DSEPTNTLAFNFDSC---EPSHDTTSNVLINISGNSISGTVGCLRP 1180
DB 1134 KMMFAQTTAGAVVAAGFC---AFDQACVCDTTOOSTSA---FNGCLAMAAPAGOKRS 1185
QY 1181 PEDELEPLDLSGPIKHNITMEM 1203
DB 1186 LPDILTFQW:APPEKSGITFDW 1208

```

```

RX MEDLINE=88179536; PubMed=2833010;
RA Gao M., Bouchev J., Curtin K., Knipe D.M.:
RI "Genetic identification of a portion of the herpes simplex virus ICP8
RT protein required for DNA-binding."
RL Virology 163:319-329(1988).
CC -1- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR. IN THE ABSENCE OF DNA REPLICATION,
CC (PREREPLICATIVE SITES). AS VIRAL DNA REPLICATION PROCEEDS, IT
CC MIGRATES TO GLOBLAR INTRANUCLEAR STRUCTURES (REPLICATION
CC COMPARTMENT(S)).
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M20165; AAA5793.1; -.
DR PIR: A28601; DMBEKS.
DR InterPro: IPR000635; Viral_DNA_bind.
DR Pfam: PF00747; Viral_DNA_bp_1.
DR K1: DNA-binding; DNA replication, zinc finger, Nuclear protein.
FT ZN: FING 499 512
FT SEQUENCE 1196 AA: 128314 MW: 0D0010A5DFA4694 Cw:64;
SO

```

Query Match 48.5%; Score 3054; DB 1; Length 1196;  
Best Local Similarity 49.8%; Pred. No. 5,3e-228;  
Matches 609; Conservative 200; Mismatches 366; Indels 48; Gaps 17;

```

QY 1 MENTOKTYV---VTPGRLGYV--ACRVEDLDLEISFLAASSTSDLLPLMRUTYE 55
DB 1 MENTOKTYV---VTPGRLGYV--ACRVEDLDLEISFLAASSTSDLLPLMRUTYE 58
QY 56 KFTSTLAVVSGARTTGAGAGITLKITSTHFPVSFVPRGCKHVPSSAANPLTFACNA 115
DB 59 SGEFANVANVVSRTTGLGTAVSLKLTSHSSSVYVHGGRHLDPSQARNTLTLCCR 118
QY 116 ABERGFSCQRPVDGAVETTGAEICTRLGLEPENTILYLVVTLAKFAVPCNYELHY 175
DB 119 ARRHGFSYTPRPDGLKHEITGEALCERLGLDPRALLYLVTBQFKKAVYCINNPLH 178
QY 176 GGLDLYHINHGVIHPLRPVYOLMPDVNRKLVDPFNTHHKSIGSEFVYPTFPYNTGLCH 235
DB 179 GGSDDVTTIGAEVHNIPVPLDLPMDPSKVTAEFPNHNKSI GENFTYPLFFNRLNR 238
QY 236 LTHDCVLAAPMAVALRVNVTAVAKGAHLAFDENHGCAYLPPIYVTVYQSSSSGTFYAR 295
DB 239 LTFEAVVGPAAVALRCRNVDAVARAAHLAPDENHGALLPADITPTAEVAGSG--KTPR 296
QY 296 GAKRNDVNSTSKPFSGGERKLASIMADTLALHEVINTGIVETPTPTIDEMPFETGM 355
DB 297 GGR-----DGGKGRPAGCGEORLASVMAAGDAALALESI VSMKAVFDPDPTDI-SAMPICBO 351
QY 356 EETPLRLNALGSTRVAVGVIGAMVFPNSALYLTVEEDSGMTEAKDGPGSFNFETQF 415
DB 352 DTAARANNAVGVYLRRAAGLVGAMVFTNSALHITVEVDAGADPRDHRK-PSFYREFLV 410
QY 416 AGPHLAANDPOTDRDGHV-----SSOSTGSSNTEFVSVDYALALICGAPLARLLFL 468
DB 411 PCTHVAANPOYVREHGVVGFEGRPAPLVLGTO-EFAGEHNLAMLCGFSFALLAKKLFYL 469
QY 469 EECDDGAFGSGHG-DALKVVTGTFSSELPSCLEKHTPRVCAMHTVYHRLQDMPFGQAT 527
DB 470 EECDDGAVYVGRQEMVYFRVADSNQOTVPCNLCTDTRACVHTLMLRKANHPKASDA 529
QY 528 RQPIGVFGTMSQYSDCDPLGNVAPYLLLRKPGDQTEAKATMODYRATLRPLERLFDLEO 587

```

```

100 540 KCAICVPCCTMSMVSMTVLICNVAAPSAIKR ADGSEIARTIMQETPAATFRMALEET 588
101 588 ERLIDRANWSSSELSSIVLHPTFRRLIDTARIEDTTQPMVLVETPRYKPELIS 647
102 648 LQYVQAVPTTAMKLETTITREALHTVNNRVQVDEVEELMKNLVGKPKFRKQV 648
103 648 PATHSMALTEPVYSCAPCTTNTFKRTHIAVVDIALISQCHVYGOVQGRNRPNQ 707
104 649 EANIMASLTLPYACGPELLOLTKRSNLAVYODALSQCHVYGOVQGRNRPNQ 708
105 708 PVLRHVRVDLNGGISTRTITVLTSEG-PVSAHPPTLGQADAPARTDGLARVSEVI 766
106 709 PVLRHVRVDLNGGISTRTITVLTSEG-PVSAHPPTLGQADAPARTDGLARVSEVI 768
107 767 RDIHVKNRVSNTNLSSEAFARVILASAYQCKRKMVLHIALFKLQFRLTFP 826
108 769 KRLVKSRLVLKAGASANSSEAKAVASLOQYOKPKRVDLLQPLFLKQFRLTFP 828
109 827 KQMPVNSKSPNQWVWLLQHNQMDIKITHEHTTTAAVKRFTVEAAINLPPTC 886
110 829 NKPVPNSQNPQWVWVWLLQHNQMDIKITHEHTTTAAVKRFTVEAAINLPPTC 888
111 887 GELAGFYVANIILKCYGHSYLINTLTSTTSGAPRPSPSVLHMPKRVTSAAITETQ 946
112 889 SELAMTANQVLLKCTHSTYHINTLTAIAGSRKRVSVQAAAM--SAQCGAGLEGA 945
113 947 KALTEKTELELWTAFTSTHLYAAMNQRPVVLGISIKYHGANNEVYVAGNNSG 1006
114 946 KALMAYAVHIEHVTSMFASNLTPVMAAPRVVLGLSTIKYHGANNEVYVAGNNSG 1005
115 1007 LMGKRNACPELIPORTKRVFLVAPRAVFAVAAVNSGSAHSSLEQDLRGIISEGA 1066
116 1006 LMGKRNACPELIPORTKRVFLVAPRAVFAVAAVNSGSAHSSLEQDLRGIISEGA 1065
117 1067 LAIYATVAVAAVAAKQIHAFFDMLSLTDEFLARDEELHQQIOTLEPTVEGAL-- 1123
118 1066 SSVAVATVKSIDPRQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 1125
119 1124 EAVKILDEKTAIGLEPTNLAFND--SCPSHDTTSNVLNIGSNGSIVSGVGLKRP 1180
120 1126 HEAVMVSQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 1173
121 1181 PEDELPELSPICRKHNTLM 1203
122 1174 FRGIDPRG EGPPIKKKGLTILM 1195

```

```

RT type 1 containing the genes for DNA polymerase and the major DNA
RT binding protein.
RL Nucleic Acids Res. 13:8143-8163(1985).
RN [3]
RP SEQUENCE OF 1062-1196 FROM N.A.
RX MEDLINE=88306232; PubMed=2457278;
RA Hamersschmidt W., Corbaths F., Mankertz J., Huh H.-J., Paul G.,
RA Ludwig H.
RT *Common epitopes of glycoprotein B map within the major DNA-binding
RT proteins of bovine herpesvirus type 2 (BHV-2) and herpes simplex
RT virus type 1 (HSV-1).
RT Virology 165:406-418(1988).
CC -1- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR. IN THE ABSENCE OF DNA REPLICATION,
CC FOUND IN THE NUCLEAR FRAMEWORK-ASSOCIATED STRUCTURES
CC (PREREPLICATIVE SITES) AS VIRAL DNA REPLICATION PROCEEDS, IT
CC MIGRATES TO GLOULAR INTRANUCLEAR STRUCTURES (REPLICATION
CC COMPARTMENTS).
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC FAMILY.
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.slb.ch/permissions/
CC or send an email to license@slb.slb.ch).
CC
CC EMBL: D10879; BA001675.1;
CC EMBL: X14112; CA42322.1;
CC EMBL: X03181; CA26940.1;
CC EMBL: M21631; AAA5787.1;
CC PIR: A03790; DNEB1.
CC PIR: B30085; B30085.
CC InterPro: IPR000635; Viral_DNA_bind.
DR Pfam: PF00747; Viral_DNA_dp_1.
DR KJ: DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
FT ZN-FING 499 512
FT SEQUENCE 1196 AA; 14844 MW; 453791625B90E9 CRC64;
Su
Query Match 48.4% Score 1049; DM 1; length 1196;
Host local Similarity 49.8%; Pred. No. 1,362,227; Indels 48; Gaps 17;
Matches 609; Conservative 199; Mismatches 367;
QY 1 MENTOKVT--VPTGPLEYV--ACRVEDIDLEISFLARSDSDALPLERNTLVE 55
DB 1 MTKKPKTATTKVPPQPLQYVYARACPSKGLR--LALISARSGSDVAVATLVGATIVE 58
QY 56 KFTSSLAIVASCARTTGLAGGITLKLTTSHFPVSVFVHGRKIVLPSSAANLTRACNA 115
DB 59 SFEENAVAVVSSKRTGGDGTAVSLKLPSSHVSSVYVHGRKIDPSDPAINTLFLER 118
QY 116 AERKQFSQDQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 175
DB 119 ARHIFGFSQVYTRPGLKHEITGEALCERLGDIDPRALLLYLVTESEKAV--INNTFLH 178
QY 176 GIDIVTHINHGDIIRIPLEPVQVLFEDVNRIVAPDPNTTHPS-TGRFVYTPFPYNTGLCH 235
DB 179 GQSDKVFICQARVHRIPVYPIQPLFMHDSKVIAPPPNANRHSIQKPKFVPIAPFPNPNR 238
QY 236 LTIIDVIAVMAVALVAVRNTAVARCAHILAFENHAGVALPDTITYTTFGSSSSSTTAR 295
DB 239 LLEFVAVGPAVALVLRNDVAVRAAHILAFENHAGVALPDTITYTTFEASQG--KTPR 296
QY 296 GARRMINVSTKPSQSGEERKIASIMADVDIAHAAVIFNCGVPEPTPIDIKEMHPIGM 355
DB 297 GGR-----DGGKGAAGGFEQVRLASVMAGDALALEISVSAVDEPPPIDIANGLFEQ 351
QY 356 EGTPLRLNALSSYARVAVGTAMVFPSPNVALYTEVENSNTKADGGPGGSENEFYQ 415
DB 352 DTAARAVAVGAVYLAHAGIAGAVVSTNSALHLEFVIDAGADPRKDHDK--DSYRHFPIV 410

```

QY	416	ASPHLAANPQDPPGCHVY-----SSQSTGSSNTEPSSVDYALICGPAALATLAFYL	466
Db	411	POTHAANPQVQDRGCHVYPGEGPPAPVAVGQV-EFAGHILAMCGSPALLAMFLYL	469
QY	469	ERTDCAFTGGHG-DALKVYTFGTTFSEIPISTICEKHTPVCATHTVHRLQRMPEFGAT	527
Db	470	ERTDCAVAVGQEMDVFRYVADSNOTDVCNLTCTPTRHACVHTTITMLIRAHHPFASAA	529
QY	528	KQPIGVECTNNQSOXSDXDPICGNAPVYLKKPKDDQDPEAKAKIMQDVRATLPRLLIDLEQ	587
Db	530	RGALVEGTNNMSKTSDDVDLQNTAAFSALKR-ADGSEIATITIQEYTRATRYVAELET	588
QY	588	ERLLDRGAPCSSEGLSSVYVDHPTFRRLIDTLKARLEQUTTOFMKVLVETRDYKIREGTS	647
Db	589	LQYVQDAVPTAMGRLEFILTNRRAIPIHVANNVQVYVDRFVQPMRMVYGRNPKFPDGLG	648
QY	648	EAHISMALTEPPYSAGCPITINELVKRTHIAVQDIALSQCHVYFGQVQVEGRNRNQVQ	707
Db	649	ENAHIMSLTDDPYACGCPPLDQLGRKSNLAVYQDIALSQCHVYFGQVQVEGRNRNQVQ	708
QY	708	PVLRKRPVYDLPNGCFILSTRSITVTLSEG-PVSAPNPTLISODAPACGTFPGDILARVSVI	766
Db	709	PVLRBRVDMENNGFELSAKTLTVALSEGAICAPSLIAGQTAPAESFEGDVARITLQFP	768
QY	767	RDIRKKNVYVSGCTMLSEARARLVGLASATQROEKRYMDHGLALGILLKQFHGLLP	826
Db	769	KELRKSVYLPAGSAAASAARAARASVLSQSAVQKPKDKRVDILGPIGLKQFHAAIPIP	828
QY	827	RGMPNPSKPPQPFMTLLQRONMPADKLTHEIITIAVRFTEYVAINFINPLEPTCI	886
Db	829	NGKPPGNSQPPQPFMTALQRLNPALRLSKEDIELIARIKKFSIDYGALINTINLAPNV	888
QY	887	GELAQFPMANILKYCHSOYLINTLITSLITGARRRDPSSVLHMRKDYTSADIEMQA	946
Db	889	SELAMYYMANQILRYCHSHSYFINLTALIAIGSRPPSVQAAMV---SAQSGALENGA	945
QY	947	KALLEKENTPELMTTAFSTHIVRAAMNRPVYVIGISISKYHGAGNNRFQAGNMSG	1006
Db	946	RALMDAVAHGAMTSMFASCNLLRPYMARAPVVLGISTKYTYGAGNDKRVFOAGNMAS	1005
QY	1007	LNGRKNVCPLETPDTRTFITIACPRGGFTCPVYGPSSGNRETTLSDQVRGIIYVSGAMVQ	1066
Db	1006	LMGRKNAPLTIPTRTKPYLACPFAGFYCAASLGGGHAESSLCEOLRGIISGGAAYA	1065
QY	1067	LAIATATVRAVGAQAMAFIDMI-STTPPEFLARLDEELHDDIIOITLETPWTEGAL-	1123
Db	1066	SSVFATVYKSLGPTQLQLLEMDLALLEDEYVLEEEMELTALALRGNNEMETDALEVA	1125
QY	1124	-EAVKILDEKTTAGDGETPTNLAFNFD--SCFESHDTSNVLINISGNSITGYVGLKRP	1180
Db	1126	HEAAALVSOLOMAGE-----VFNGDPOCE-----DDNATPPCGPGAPGAFAGRRKA	1173
QY	1181	PEDEDLFDLSIGPIKHGNTMEM	1203
Db	1174	FHGSDPFG-EGPPQKKGDLTDM	1195
RESULT 5			
DNBI_HSVIE			
AC	P17469;	STANDARD;	PRT; 1196 AA.
DT	01-AUG-1990 (Ref. 15, Created)		
DT	01-AUG-1990 (Ref. 15, Last sequence update)		
DT	16-OCT-2001 (Ref. 40, Last annotation update)		
DE	Major DNA-binding protein (infected cell protein 8) (ICP 8 protein).		
GN	DBP OR UL29 OR ICP8		
OS	Herpes simplex virus (type 1 / strain F)		
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;		
OX	Alphaherpesvirinae; Simplexvirus.		
NCBI_TaxId=10304;			
RN	[1]		
RP	SEQUENCE FROM N.A.		

RX MEDLINE:88306231; PubMed:2841193;  
 RA Hammerschmidt W., Conraths F., Mankertz J., Pauli G., Ludwig H.,  
 RT Bubb H. J.;  
 RT "Conservation of a gene cluster including glycoprotein B in bovine  
 RT herpesvirus type 2 (BHV-2) and herpes simplex virus type 1 (HSV-1)."  
 RL Virology 165:388-405(1988).  
 CC -I. FUNCTION: SINGLE STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA  
 CC REPLICATION.  
 CC -I. SUBCELLULAR LOCATION: NUCLEAR. IN THE ABSENCE OF DNA REPLICATION  
 CC FOUND IN THE NUCLEAR FRAMEWORK-ASSOCIATED STRUCTURES  
 CC (PERREPLICATIVE SITES). AS VIRAL DNA REPLICATION PROCEEDS, IT  
 CC MIGRATES TO GLOBULAR INTRANUCLEAR STRUCTURES (REPLICATION  
 CC COMPARTMENTS).  
 CC -I. SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN  
 CC FAMILY.  
 CC PIR: D29242; DMBEHF.  
 DR InterPro: IPR000635; Viral\_DNA\_bind.  
 DR Pfam: PF00747; Viral\_DNA\_bp\_1.  
 KW DNA-binding; DNA replication; Zinc-finger; Nuclear protein.  
 FT ZN\_FING 499 512 C2HC-TYPE  
 SO SEQUENCE 1196 AA; 128373 MW; BC672584DDBIC8E2 CRC64;

Query Match	48.48;	Score 3047,	DB 1,	Length 1196;
Best Local Similarity	49.88;	Pred. No. 1.8e-227;		
Matches 609;	Conservative 199;	Mismatches 367;	Indels 48;	Gaps 17

QY	1	MENTOKVTV---VPTGPGYGY--ACREJEDJLEISLARSTSDIALPLMRNI	55
Dp	1	METKPKIATITLKVPPGPIGYYARACPSGEGIEL--LALLSARSDSDYAAVPLVGLTIVE	58
QY	56	KTFSSSLAVSGARTTGLAGAGTILKLTTHFSPSVFVFIHGKHLVDPSSAPNLTIFACMA	115
Dp	59	SGFENAVAVVSGSKRTJGGTAVSLIKLTPSHYSSVYVFGHGRHLDPSTOAPNLTCLER	118
QY	116	ARERGFSGCGPPVDGAVETTGAEICTRLDEPENTLILYVVALPKFAFMCNVLIAY	175
Dp	119	ARRHGFEDTYPFRGDLKHETTGALCERLDGDPDRALLIYVTEGKREANISINNTLEHL	178
QY	176	GGDLIVHNGDVIRIPLFVQULFMPDVNRLVPPDFNTHHKSIGEGHYPTFPYNTGCH	235
Dp	179	GGSKVITIGGAETHIRIPIYPLQTLMPDPSRYIAELFPNANHRSIGEKFTYPLFPFNRLNR	238
QY	236	LIIHCVIAPFMAVVALRVRVNTVAAGAHLADENHEGAVLPEDLIITYFOSSSGSTTAR	295
Dp	239	LIFEVYVGPAAVVALRCRNVDAVAAAHAHLADENHEGALPADIPTAFAPESQG--KTPR	298
QY	296	GARRNDVNSISKPSPSGGEFERRLASIMAADPLAAEVLPNGVLEEIPPTDITKEPMPLIGM	355
Dp	297	GER-----DGGGKGAGGFEGRRLASVMAAGDAALLLESIVSNAVDEPPTDISAMPLEEGQ	358
QY	356	EGTLPRLNALGSTYARVAGVIGAAVVSFNLSALLYLFEVHUSMTFAKDCGPPSNRRYQF	415
Dp	352	DTAARAAVAVAYLAKAAGLGAVAFSINSALHLEVDVDAAPADPKDHSK--PSYKFFPLV	412
QY	416	AGPHLAAMPQTRDQGHVY-----SSQSTSSMTPEFSVDLITCGGAPILARLLLEFL	466
Dp	411	PCHVHAANPOVDRGRGHVPPREGPTAPLVGCTO--ETAGEHIAHLCGSHALLAKMLPYL	463
QY	469	ERCDAGAFTCGHG--DALKYVTGTFDSEIPCSLCERKHTRPVCAHTVHRLKORMRPOGAT	522
Dp	470	ERCAGVAVIGROEMDVRYVADSNQOTVPCMLCTFDFPHAVHTTLPMPARHPKFSAA	525
QY	528	KOPIGVGTIMNSOYSDDDPGLGNVAPYLLLRKPGIOTFAAKTMDVYKATILIRIPLTILQ	585
Dp	530	RGALIGVGTIMNSMTSDCVLGNTVAFSALKR--ADGSEFARTIMQETYRAATERMAELIET	588
QY	588	ERLLDRGAPCSSEBLSIVYDHFERRLLDLTARIELTTUJEKVVLEVETRDYIREGLS	645
Dp	589	LQYVDQAVPTAMGRLELITLREALHTVYVNNVNOVDRBHVQILMRNLVDEGNPEFFRQICG	648
QY	648	EATISMALTEPYSGACAPLITNPLVKRTHLAVVDLALISUCHCVYVQVQVHGKMPKNOPO	707
Dp	649	EANIAMSLITDYPYCGCPRLDILGRSNTLAVVDLALISCHHGFAASQVSEGRNFRMFO	708



DB 952 LDAHPCAMTSMFASCNLKPVMAARPMVVLGSLSTSKYYGMAGDRPVQACNNASLIGCKN 1011  
 QY 1013 VCPLETFDTRERFIACPRGGFICPVTPGSSGNRETTLSQVKGITVSGAMVOLAIVAT 1072  
 DB 1012 ACPLIFDTRKRFVLAACRAGFVCAASSIAGGAAHSHSLCEFOJLKGIIAEGAAVASSVFA 1071  
 QY 1073 VYRAGARAOHMAFPDMLSLTDEFLARDLELHQIOTLETPMTVEGAL----FAVKI 1128  
 DB 1072 TVKSLGPRTOQOQIEDMLALLEDEYLSSEMEFTTRALRGHGEKSTDALEVAHEAEAL 1131  
 QY 1129 LDEKTTAGDGETPTNLAFNPFDSCEPSHDTTSNVLNIGSNISGCTVPGLKRPEDDELDP 1188  
 DB 1132 VSQICAAAGH-----VFNFGDFGDHDDHAASFGLAA--AGAAVAKKRAFHGDPPG 1182  
 QY 1189 LSGPIKIGNITMEM 1203  
 DB 1183 -EGPPEKK-DITLDM 1195

## RESULT 7

DNBL\_HSV2 STANDARD: PRT: 1197 AA.  
 ID DNBL\_HSV2 STANDARD: PRT: 1197 AA.  
 AC P36384:  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Major DNA-binding protein (infected cell protein 8) (ICP 8 protein).  
 GN DBP OR UL29 OR ICP8.  
 OS Herpes simplex virus (type 2).  
 OS Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.  
 OX NCBI\_TaxID=10310.  
 RN [1]  
 RP SEQUENCE FROM N. A.  
 RX MEDLINE=93228441; PubMed=8385914;  
 RA Toh Y., Liu Y., Tanaka S., Mori R.;  
 PT "Nucleotide sequence of the major DNA-binding protein gene of herpes  
 RT simplex virus type 2 and a comparison with the type 1.";  
 RL Arch. Virol. 129:183-196(1993).  
 CC -1- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA  
 CC REPLICATION.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN  
 CC FAMILY.  
 DR InterPro: IPR000635; Viral\_DNA\_bind.  
 DR Pfam: PF00747; Viral\_DNA\_bp; 1.  
 DR DNA-binding: DNA replication; Zinc-finger; Nuclear protein.  
 FT ZN\_FING 499 512 C2HC-TYPE.  
 FT DOMAIN 1197 1197 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT SEQUENCE 1197 AA: 128412 MW: 157665865DBF CRC64:

Query Match 47.9% Score 3013 DB 1: Length 1197;  
 Host Local Similarity 49.0% Prev. 7.9% 2.5,  
 Matches 549; Conservative 213; Mismatches 367, Indels 40; Gaps 16;

QY 4 TQKTVPTGFLGVV--ACREDEDELEISFLAARSTDSLALPLLRNTLVEFTSS 61  
 DB 7 TTTTIVKVPQMGVYVYGCAPARGLLEI--LSLSARSDDAIVAAPIVLGLTVSSGFAN 64  
 QY 62 LAVVSGARTTGLACAGITLKTISHYPSYVVFHCGKHLVPSSAAPNLTRACNAARFPG 121  
 DB 65 VAAVVGSGITGLGTAVSLKLPMSHSPSYVVFHGRILAVSTQAPNLTRCERARRHF 124  
 QY 122 FSRQCGPPVAVCAVETGAEICTRIGLEPENTILVLTALPKFAVFNQNVLAHGGIDIV 181  
 DB 125 FSDVAPRCDLKHETTDALCHRLGIDPDRALLVLTITGCRFAVNCISNFTLHGGMKV 184  
 QY 182 HINIGVIRLPLEFVQLMPDVNRLVDPENTIHRSIGEGVYPTPYNTGLCHLHDCV 241  
 DB 185 TIGDAEVHRIPIVYPLDMHDPFSKVIADPFNCNHRISIGENFNYPDLFFNRLARLLEAV 244

QY 242 IADMAVALRYNTAVARGAHLAFDENHCAVLPDITVTFQSSSSGTTAKGARND 301  
 DB 245 VGPAAVALRARNDVAVARAHAHLAFDENHGAALPDITITAEASG--KPRGAR-- 299  
 QY 302 VNSTKPSPSGGERLASSIMADTALHAVENTIGYEETPPDIKEMPEICMEGTLPR 361  
 DB 300 --DAGKKGPRGGEORIKLASMACDAALESIVSMVAFDPPDITWPLLEQGFYPAAR 357  
 QY 362 LNALGSYARVAVGIVAMVFPNSALYTEVEDSGMTEAKDGPSPFNFFYOFAPSHLA 421  
 DB 358 AGAVGAYLARAAAGLVCAWVFSTNSALHLEVEDAGADPRDHSK--PSFYFTLVPGTHVA 416  
 QY 422 ANQTDRODGHV-----SSOSTGSSNTEFSDVYALICFGAPILARLLFYLERDAG 474  
 DB 417 ANQDREGHVVPGEGRPTAPLVGGO--EFAGEHLAMLCGFSPALLAKMLFYLERDGG 475  
 QY 475 AFTGGH-DALKYVGTGTFPSEIQCCEKTRVCAGHTVHRLROMPRGUAIKPILV 533  
 DB 476 VIVGQEMDVFRVADSGQDVPNCICTFETRHACAHITLMRLRAHPKASARAGIGV 535  
 QY 534 FGTMSQYSDCDPLGAYAYLLIRKPDOTEAKKATMDOTYPATLEPLEFLEQERLDR 593  
 DB 536 FGTMSAYSDCVLGNVAFSALKR ADGSENTRTIMOETIRATERVMALEALQYVD 594  
 QY 594 GAPCSSEGLSVIVDPTRRLIDTLARLEQTTQFMKVIYETRYKI REGISEATHSM 653  
 DB 595 AVPTALGRLTEITIGNEALHTVNNIKOLVREVEDMLNLEGRNFKFRDGLAEANHM 654  
 QY 654 ALTFDPSGAFCDPTNLELKRTHLAVVODLALSQCVEYGVVESEKRFNFGUPLARR 713  
 DB 655 SLIDPYTCGPCLLOLLARRSNLAVYODLALSQCVEYAGVAGVBCGRNFRNGQVPLARR 714  
 QY 714 FYDLEFNGFISTRTITVLTSEG--PVSAHPITIGQDAPARTDGLARSEVETIRIK 772  
 DB 715 VMDITNNGFLSKATITVALSEGAICAPSLITAGQTPAASSSEGDVARTLTFPKELRX 774  
 QY 773 NRVEFSGNCTNLSEARARIVGLASAYORQKRVLMHAGLFGILKOPGLLFPKPMNP 832  
 DB 775 SHVLFAAGASANAEEAKKAVASLOSAYQKPKRVLDLPLGFLKQFHAVIFPNKPLPG 834  
 QY 833 SKSPDQWFTLLOHONMPADKLTHEITTAIVKFTETEVAINFINLPPTICGLAOF 892  
 DB 835 SNQPNQWFTLLOHONMPADKLTHEITTAIVKFTETEVAINFINLPPTICGLAOF 894  
 QY 893 YMANLILKYCHSOYLINTLITSIITGARPRDPSSVLIHMRDVTSAADIEFOAKALEX 952  
 DB 895 YMANOILRYCHDSTYFINTLTAVIGSRPSPVQAAANAPQ---GAGLEAGARLMD 951  
 QY 953 TENLPELMTTATSTHILVRAMNORPYYVLGISISKYHCAAGNNVPPQAGNMGCKN 1012  
 DB 952 LDAHPCAMTSMFASCNLKPVMAARPMVVLGSLSTSKYYGMAGDRPVQACNNASLIGCKN 1011  
 QY 1013 VCPLETFDTRERFIACPRGGFICPVTPGSSGNRETTLSQVKGITVSGAMVOLAIVAT 1072  
 DB 1012 ACPLIFDTRKRFVLAACRAGFVCAASSIAGGAAHSHSLCEFOJLKGIIAEGAAVASSVFA 1071  
 QY 1073 VYRAGARAOHMAFPDMLSLTDEFLARDLELHQIOTLETPMTVEGAL----FAVKI 1128  
 DB 1072 TVKSLGPRTOQOQIEDMLALLEDEYLSSEMEFTTRALRGHGEKSTDALEVAHEAEAL 1131  
 QY 1129 LDEKTTAGDGETPTNLAFNPFDSCEPSHDTTSNVLNIGSNISGCTVPGLKRPEDDELDP 1188  
 DB 1132 VSQICAAAGH-----VFNFGDFGDHDDHAASFGLAA--AGAAVAKKRAFHGDPPG 1182  
 QY 1189 LSGPIKIGNITMEM 1203  
 DB 1184 -EGPPEKK-DITLDM 1195

## RESULT 8

DNBL\_HSV2 STANDARD: PRT: 1186 AA.  
 ID DNBL\_HSV2 STANDARD: PRT: 1186 AA.  
 AC P12639:





Query	Subject	Score	Length	Match	Gap	Ident	Score	Length	Match	Gap	Ident
13	GELGVYVACREDDLEISFLARSDSDLLPLMNLIVERTFTSLAVVSCARTTG	72	1128	23.2%	186	537	669	1128	23.2%	186	537
18	GPGCGIYFYPATYPLRLEAVATLGTGAGKCLTYPLLGITVEPFSIN	74	1128	23.2%	186	537	669	1128	23.2%	186	537
73	LACAGITLTKTTSHPYEVFVFGHGKHYLPSSAAPNLTRACNAARERGPS	126	1128	23.2%	186	537	669	1128	23.2%	186	537
75	DPNGL--LNAVSYHRIYVFNHNMWPRFPEGGLGALCGETREVGIDATYALPRES	131	1128	23.2%	186	537	669	1128	23.2%	186	537
127	GRPDVAGVETTGAEICTLCTGLGEPENTILYLVYVALFKNAVTCMVNFELHGGDLVHINHG	186	1128	23.2%	186	537	669	1128	23.2%	186	537
132	SKPGDFPE-----GLDPSAYLCAVAVATFEAFKRLISGNI.VAIPSLKQEVAGOS	181	1128	23.2%	186	537	669	1128	23.2%	186	537
187	DVIRLPEFVOLLFMDPVRLVLPDPFNTHRSIGEFVYPRPFYNTGLCHLHDCVIAMA	246	1128	23.2%	186	537	669	1128	23.2%	186	537
182	ASVAVPELTKVEFVBQVL-----KQFYNSDLSRCMHEALYTLA	222	1128	23.2%	186	537	669	1128	23.2%	186	537
247	VALRVNVTAVARGAALHAFDENHEGAVLPDITVYTFQSSSSGCTTYARCARNDVNST	306	1128	23.2%	186	537	669	1128	23.2%	186	537
223	QALVRYRRYKGL-----VELLEKOSJODAKAVKAPLKEFPAST	261	1128	23.2%	186	537	669	1128	23.2%	186	537
307	KPSPSGGERLTLASIMADI--ALHAEVIFNTGIYE--EPTLDIK--EMPAFICMEGLP	360	1128	23.2%	186	537	669	1128	23.2%	186	537
262	LSHDSG-----ALMIVDSACELASYAPAMLASHETPASIINAYDSWPLPAIDCEPFA	315	1128	23.2%	186	537	669	1128	23.2%	186	537
361	RLNALGSTARVAGVIGAMVSPNSALYLTVEVDSGMLEAKDGGPGSPFNFFYOFAG--	417	1128	23.2%	186	537	669	1128	23.2%	186	537
316	RVALHLHRYNASTLAPVSTQIPATNSVLYV-----SGVSKSTGGCKESLEFNFTMHG	370	1128	23.2%	186	537	669	1128	23.2%	186	537
418	-----PHLAANQTDROGHVLSQSGSSNTFFSVDTYALICGFCAPILARILF	466	1128	23.2%	186	537	669	1128	23.2%	186	537
371	LOGSTWDCRRPCESGWGQDPVVG-----NGPAIN--VAVEHLYVAAVSFNPILARVY	422	1128	23.2%	186	537	669	1128	23.2%	186	537



```

QY 356 EGTPLRLNALGSTARAVGIGAMVFPNSALYIEVEDSGMTEAKDGGPSENFYOF 415
DB 315 FTFEDRHKAIQIMNAMSVHYTHLPSTNSVLTLLKINMOTOSKSEON---YNYNYFMQ 371
QY 416 AGPHIAANQOTDRDGVIVSSQSGSNTESVYUVALICQFGAPLLARLLFYLERCDAGA 475
DB 372 HCLVAADATQURENDEBPASGAFKFSGGTYTLTHLALASSFPHLLAKNYYMQFQHQK 431
QY 476 FTGCHQDIALKYVYGFDSHPCSLCEKHTKRPVCAHTTHVRLRQRMDR-PCQATROPICGV 534
DB 432 STTNANYSVPQVYGTAAADICLCOGTCRASCITHTLFLRLKRPVPLVLSQSRDPVYV 491
QY 535 GTMNSQYSDCDPLGNAYAPLLLRKPGDQTEAAKATMODTYRATLEBLFDLEQEPILDPG 594
DB 492 C-VSGQYNDLDMIGNAFTE---KKKDEAVQNAESEKTYTMOIQWV---EKLSTMG 542
QY 595 ACCSSGGLSSVIVDHPTRHLLDTLAKRLQOTTQPMKVIYERDVKIRGLSEATHSMA 654
DB 543 VREGTGV-SELTIDIOSLKTFRDIDNVDSYVKFMCLVK-NNINRETTIKTVHVLH 600
QY 655 LTFDPYSGAFCDITNPLVNRHLAVVQDLAL-----SQCHCFYGOOV 697
DB 601 YCCNWFQWAPCAMPLINI.FYKSVLAIIDIDICLPLAMTYEQDNPSICMMPSEMLKVH--QT 658
QY 698 EGRNR-----NQPQVLRKRFVDLRNGGISTRTITVTISEGVSAPNFTLGQ 746
DB 659 IWTNRKSSCLDRQVLTGSHKIVHTDMPDFLN-----IDSALS-----CQ 699
QY 747 DAPAGKTFMDLAVHSEVYIRDIRVKNRVVFSQNTNLSEAAARLVGLASAYQROEKV 806
DB 700 IVP--MKQVRLAKALLTPKTIKIKRIVFSN--SSMETIIGSGIKSAT-----KKS 750
QY 807 DMJHAIQILKOPHGLLEPRGMPSKSPNPQWFTLLQRMQPA-DKLTHERETITIA- 864
DB 751 YLVTPYKMFITSLKRVMP-----PNAKISALYLMHTFESQKQOLPVLPCISRENMVELAN 805
QY 865 AVKRTTEEAALNFNLPTCIGELAEUFMANLILKYDHSQYLINTLSI-----ITG 918
DB 806 YVETSSKHHMDMNVLDIIPITILLYAKVANNITLKTGCGOTQVATTLQCLLPITQFTISA 865
QY 919 AKRPR---DPS---SVLHWIR-KD-----VTSADIEFQAKALLEKTENLPELWTFAT 964
DB 866 TEYPIVLLDQSSIMSYDHLISIKDKHALTVQTLKEDATVK-----908
QY 965 ISTHLYKAAAMNRPVAVICISISKYHGAAGNKNRVQAGN---WSGINGCKNVCPLFT-ND 1020
DB 909 -----QRPVTVHVLVANKYTGINGNQIFQCCMLGYFMGQGVDRNLIPDSTGR 957
QY 1021 RTREFIACPRGGLCP---VTGSSSGNRET---TLDQVRGIIIVSGAMVQALAIYA 1071
DB 958 KQNNSSYKRRHNVHTPMVAHLVKKNSNINNTPEVETIRKKNVONIFELKD---NINIFD 1014
QY 1072 IV---VAAVCARQAHMAHMDMISITDHEFLARHIFELHNDITITQITETPWTEGALFAV 1126
DB 1015 NVLELVKGLDSCENITEDDQFYLGEEYIMSDEIWSRFQITDTSGAWSEVNTKVL 1073

```

```

RA Gompels U.A., Nicholas J., Lawrence G., Jones M., Thomson B.J.,
RA Martin M.E., Estathiou S., Craxton M., Mccaulay H.A.;
RT "The DNA sequence of human herpesvirus-6: structure, coding content,
RT and genome evolution."
RL Virology 209:29-51(1995).
CC -1- FUNCTION: SINGLE STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sdb.ch).
CC
CC EMBL: X92436; CAA63167.1; -
CC DR EMBL: X83413; CAA58375.1; -
CC DR InterPro: IPR000635; Viral_DNA_bind
CC Pfam: PF00747; Viral_DNA_bp.1.
CC KW DNA binding; DNA replication; Zinc-finger; Nuclear protein.
CC FT ZN_FING 459 475
CC SQ SEQUENCE 1132 AA; 127762 MW; 4468D3E559F02D8 CRC64;

```

Query Match 7.8%; Score 492.5; Dh 1; Length 1132;  
Best Local Similarity 21.8%; Pred. No. 1,2e-29;  
Matches 274; Conservative 185; Mismatches 544; Indels 253; Gaps 53;

```

QY 2 ENTQKTVYPTGPGYVACRVEDLDEISFLAARSTDSDLLPLMRILYEKTTSS 61
DB 3 DENEIVSAPVSTAAMLTYPPKDELDVLSLMEKRNPIYISPLMLTYENDSST 62
QY 62 LAVVSGARTITGLAGATITLKTTSHEPPSVFVTHGKHLV-PSSAANLTRACNAERF 120
DB 63 V-----RIPITNFGIILTKITSEMPVCFPHJTEGLVGMEDHSDLLRLCEURK 115
QY 121 GFSRCQPPVDGAVETGAEICRLEPENTILYLVVIALFPAVPMCVPLHYGGLD 180
DB 116 HDSFEVPTARKYIDKA--LCSAVGKADSDVICHVACNGFEELF-----AGLLI 165
QY 181 -----VHINHGDIPIPLFVQDLPMDVNPVLDPPENTHRSIGEGVYPTPYNTG 232
DB 166 PCVEEQIQVQVGVSCVKKILYSATILETE-----ETISLSSCTEFLQKCFPIA 216
QY 233 LGLLHNDYIAPAAVALRPNYTAVAAPGAHILAFENHNPVAVLPDITTYFOSSSGTT 292
DB 217 LSETLTYVFTSKGTLIRSNTELIDAGLKQFTQDGEVTVKLAPRKYLV----- 266
QY 293 TARGARRNVNSTSKPSPGCFERRILASIMAAPTALHAEVFN-----TGVEFTPTDI 346
DB 267 ---GISQKISAVEK-----FLMLVDQV-TELSFHVAYELISVVD--PSQI 309
QY 347 ---KEMPMFIGMEGTLPRLNALGSTARAVGIGAMVFPNSALYIEVE-----D 394
DB 310 MNFNDMPILIRNSETTAERMAOLTNILKHLSSHILAVLIFAPNSILYOSKLAFLIPNVQAFN 369
QY 395 SMTTEAKDGGPSENFYOFAGPHLAANDQUTDRDGH-VLSQSGSNTESVYUVALI 453
DB 370 SWVTO-----ELLRLSLFNCALSSLDVYNNDRKTIKICSTSGKDDKFSANHLAYA 422
QY 454 CGFGAPLLARLLFYLERCDAGAFGCHGALKYVGTGTFUSEIPCSLCEKHTRPVCAHTTV 513
DB 423 CATSPQLSFEVYVNNLRMSVYNAHNAHTEIYHNLVNC--SANLCEPQDCKCCQSGITAM 480
QY 514 HRLQEMRPFQATROPICGVFTMNSQYSDCDPLGNAYAPLLLRKPGDQTEAAKATMOT 573
DB 481 VAVGTRLPALIPKNNVKEPLVMSMFSRYTAENVLDLGSFG-----RKTVSELKEIGKDOON 535
QY 574 YRATLERFLDDEQLDRGA-----PCSSGGLSSVIVDHPTR-----RIIDT 618

```

```

146 546 LS ----- LDREKESVJIFVYFKKNSLIEPVTSEETFNVSCKKLEFVSIH 580
147 619 LKARLEQTTLOPMKYLVEREYKI-REGISSEATHSMALTFDYSAPFCITNPLVKRTHL 677
148 581 LKCTLEKESV--KCTVEKKRQUTTHREJENLOSNNVITTYATAFSPFLFSTYKVL 637
149 678 AVVDALASOCHVYGVGVGFSENFENFGVPIPPPEVDF-----NAGEFISTRIT 729
150 648 IYLOMAL-IVASGHVVDKRTGNSISKMLVOYQSLYGFHSSYLAKKFLNTRIVK 693
151 740 VLSSEPPASAPNPLIGQ-DA-----PAGK-PPGIDLARVVEVRIKRVNRYVSGN 780
152 694 V-----ASVNDMEQLDLDLYKSGKRYKTTIQAKTFLSMGLDLDERIKNPENKSS 745
153 781 CTNLSSEARARIVGLASAVORERKVDMLHGAIGFLKQFHGLFPRCPMPNSK-SPNQ 839
154 746 KFAHNNPYFKKV-----KHAANNLSGCTSLFKKRYHKKL-----PAVKISGLFL 791
155 840 MEWILLORNMVAKLTIEHTT-IAAVKRTVEEYVAINEFNLPTGIGELAOFYMANLI 898
156 792 WQRFLLNNVPKLLDQNDPEVKTFIKFASTINTEYDEIDIDIDPECLSTFLDCYFHNKF 851
157 899 LKCTHSGYL-INTLS-ITIG-----ARRPRDS--SVLHMLKQVTSADIEET 944
158 852 LSALEPHDITSLRDLTSKLVTONFVLPVLDKQPKRSSIOEYLVYKVLDDG--VFN 909
159 945 QAKALTEKENLPELMTAFSTHVRAMNORPVVGLISTISKYGAAGNRRVFOAG-- 1002
160 910 FVLASIKRPN-----FGLHETS-----KSLVFLGLLEKFLVSLA-NHRYFOFQGL 954
161 1003 NMSGLNG-GKNVCP-----FTEDTRKFLIACPRGGFICVYTGSSNRETTLSQ-- 1053
162 954 GMFGSGVDRNINIPSSALQDPRFM-KDTLATKPSVIV-----KKYRKALIMPTREV 1008
163 1054 VAGTIVS-----GQAMVOLAIVAVRVARGAQOHMAFDMLSLDD-ETLARD----- 1102
164 1009 VKGKALSTVENITNDIDPELLIAEVM--DREDKPTMDQLFFVDGSEALASIMKL 1065
165 1103 EELHVVITLITLTFWFGALPKAV-----KILDEKTTAGD-----ETPT 1142
166 1066 NMLVIMVKNQDSEI-NQSVFPEVSSNAPVVDSELLAEEDQGNQVLCDETEET 1120

```

```

RESULT 13
167 DMBI HSV6Z
168 AC P52548;
169 ID DMBI HSV6Z STANDARD PROT 1132 AA.
170 01 OCT 1996 (Rel. 34, created)
171 01 OCT 1996 (Rel. 34, last sequence update)
172 16 OCT 2001 (Rel. 40, last annotation update)
173 Major DNA binding protein (MMP).
174 041 OK KAL.
175 Human herpesvirus (type 6 / strain 729) (HHV6).
176 Viruses, dsDNA viruses, no RNA stage, Herpesviridae;
177 Herpesvirales; Kaposi's sarcoma-associated herpesvirus.
178 NCBI TaxID:36351;
179 RN 111
180 SEQUENCE FROM N.A.
181 MEDLINE 95074921; PubMed-7984761;
182 Stanley F.R., Damjanovic G., Black J.B., Damhaugh T.R., Pellet P.E.;
183 "Molecular cloning and characterization of human herpesvirus 6B orf1y1
184 suggests acquisition of orf1y1 by transposition";
185 J. Virol. 69:589-596(1995).
186 FUNCTION: SINGLE STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
187 REPLICATION.
188 SUBCELLULAR LOCATION: Nuclear (Probable).
189 SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
190 FAMILY.

```

This SwissProt entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

```

191 modified and this statement is not removed. Usage by and for commercial
192 entities requires a license agreement. (See http://www.isb-sib.ch/announce/
193 or send an email to license@isb-sib.ch.)
194 CC -----
195 DR EMBL; A157706; AAB06339.1; -
196 DR InterPro; IPR000635; Viral_DNA_bind.
197 DR Pfam; PF00747; Viral_DNA_bp.1.
198 DR DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
199 FT ZN.FING 459 475
200 SO SEQUENCE 1132 AA; 127817 MW; 839AB46A5104EFP CRC64;
201 Query Match 7.8%; Score 490.5; DB 1; Length 1132;
202 Best Local Similarity 22.0%; Pred. No. 1.7e-29;
203 Matches 278; Conservative 180; Mismatches 535; Indels 271; Gaps 54;
204 2 EMTQKIVYPTGCPGLGVYACRVEDLLEHLSPLAASSTSDIALPLMNLVYKPTFS 61
205 3 DENETVSAAPVSTAAWIVFPEKELDLVSLTSMERSPVVISPLMLNLTVENDEFT 62
206 62 LAVVSGARTGLAGAGITLKTLSHFYPSVFEHGGKHYL-PSAPNLTPLNMAREFE 120
207 63 V-----KTPITNCGTLLTKITSMVPCFPHGTHQVCAAHHDGDLIRICRQKRF 115
208 121 GFSKCGPVDGAVETTGAEICTRLGLEPENTILYLVTALKEAVYMGVFLHYGDI 180
209 116 HUSEVEVPAPFVIDIK-LCSAVGKQASVLIHVAQNGSEKELLE-----AGLLI 165
210 181 -----VHINHGDIKPLFPVOLFMDDVNMKLVDPNTHRSIGRQVYPTPYNG 232
211 166 PVEFEOYQVGEYSVKPLYSATIFETE-----ETLSLSTGTEFIDEGFLPA 216
212 217 LSETLFFYVFTSMGTLIRNSNTEKELIDAGIKQFQGGQVKAIPKHYI----- 266
213 233 LCHLIDCVIAPMAVALPVNVTAVAPGAHLAFDENHGAVALPPDITVYFGSSSGTT 292
214 217 LSETLFFYVFTSMGTLIRNSNTEKELIDAGIKQFQGGQVKAIPKHYI----- 266
215 293 TARGARNDVNSTKSPSGRFRPLASIMADPTALHAIVN-----TGIEETPTDI 346
216 267 ---GISGKISAVEK-----FLMLVDSSV-TELSFSHVAEYLDSSVD-PSGI 309
217 347 ---KMPPTGMRGCTIPRINALGSYARVAGVIGAMFSPNSALVTEVE-----D 394
218 310 MNENMPITIRNSETHAERMAQLTNLKHLSSHILAVILFAPNSILYCSKLAFIPNVQAN 369
219 395 SGMTEAKKGQSPSPKRFQVGFAGPHLAAMPQIDKQH-VLSQSSTSSSTERSVDYALI 453
220 370 SVMTO-----ELLKRSISPCNALSSILREYVNDKRLIKDSYSGDKDKESANHAYAA 422
221 454 CGEGAPILATLLEFLERCAGAFGTGCHGDALKYVTGFPESEI-----PGLFEKH 503
222 423 CATGGLLSYVVMNLMKSV-----YNAINATELYNHLVNSANLCEETDCK 470
223 504 TRPVCAHTTVHRLKQMPHFGATIKQIPGVPTMNSOVSDDCHLGNVAVYLILKQIGDT 563
224 471 CQSGTGTAMVAVGTPPLPAIPKNVKEPELVMSNFSYVAEVLIVSFG-----FKVSEL 525
225 564 EAAKATMODTYRATLERLFDLEQELLDRGA-----PSSSEGLSSIVDHPTR- 613
226 526 KETGKQDQNTLS-----LDKGFVVSQIFDYCKKNSLIDVVTGSHIVNVS 570
227 614 -----PILDTLARIQTTTQPMKYLVEREYKI-REGISEATHSMALTFDYSAPFC 667
228 571 KKDPSILHGLUGLEHVS--KCTVEKKRQUTTHREJENLOSNNVITTYATAFSPFL 627
229 668 TNPVAKRTHLAAVQDLASOCHGVFYGVGVGFSENFENFGVPIPPPEVDF-----N 719
230 628 LTFAYAYKVLITVLOMAL-----IVASGHVVDKRTGNSISKMLVOYQSLYGFHSS 683
231 720 GGFISTRITVLSKSPVAPNPLIGQ-DA-----PAGK-PPGIDLARVVEVRIKRV 770
232 684 KGLNTRIVK-----ASVNDMEQLDLDLYKSGKRYKTTIQAKTFLSMGLDLDERIK 735
233 771 VKNRVFSGCNTLSSEMAPRLVGLASAVORERKVDMLHGAIGFLKQFHGLFPRCPMP 830

```

```

Db 736 IKRPNKSSKAHNNPFEKKNV-----KHKKNPLSGCISFLFKYHDKLF----- 781
QY 831 PMSK-SPNQWMTLLQBNMPPADKLTHERIT-IAAVKFTEEYAININLPTCIGE 888
Db 782 PNVKISCLEHMOBELLNNPKLIDIGNPEEVKFKIFAFSITNTYDEIDIDIQPELST 841
QY 889 LAQFVAMNLIKCYDHSOYL--INTLTS-1ITG-----ARRPRDS--SVLHWIRK 934
Db 842 FIDCYHNNKFLASGHFDLITSLHGLTSKLYTONPVLFPVLDKQKPFSSIOEYLYVKK 901
QY 935 DVTSAADITQAKALLEKTEENDELMTTAFSTHVLVRAANRPVVVLGISISKYHGAAG 994
Db 902 LVLDG--VNPVPAJASISKEPN---POTITFS-----RSLVTFGLTLEKPVSLA- 944
QY 995 NNRVEQAG--NMSGNG-KGNVCPPL-----PTEDFTRRPIIACPPSGGFCPTGSSGN 1045
Db 945 NNEYFOGOLGWTGGSGVDRNLNPTSSALODFEFR-OKTITATKSEVIV-----KVR 998
QY 1046 KETTSISQ--VWQIIVS-----GGAMVCLAIYAIVKAVGARAQHMADMLTDD-E 1096
Db 999 RETIMEDLEVKGKLSIYENLTNDIDPELLITAEVVR--DREDKPTMDMLFVYDRE 1055
QY 1097 FLARD-----LEELHDDIQTLETPWVEGLEAV-----KILDEKTTAGDS--- 1138
Db 1056 ALAASIMLKINHLDVNNVPSIA-NIQSVFAVSSNDAPVYDFSHLLAEEDDQASGLK 1114
QY 1139 --ETPT 1142
Db 1115 CDETER 1120

RESULT 14
DNBI_HSV7J STANDARD: PRT; 1131 AA.
ID DNBI_HSV7J
AC P52339;
DT 01-OCT-1996 (rel. 34, last sequence update)
DT 01-OCT-1996 (rel. 34, last sequence update)
DT 16-OCT-2001 (rel. 40, last annotation update)
DE Major DNA-binding protein (MDBP)
GN 041.
OS Human herpesvirus (type 7 / strain J1) (HHV7).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxId=57278;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=J1;
RA Nicholas J.;
RC Submitted (JAN-1996) to the EMBL/Genbank/DDJ databases.
CC -! FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION.
CC -! SUBCELLULAR LOCATION: Nuclear (Probable).
CC -! SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U43400; AAC54703;
DR InterPro: IPR000635; Viral_DNA_bind.
DR Pfam: PF00747; viral_DNA_bp; 1.
KM DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
FT ZN-FING 459 475 C2HC-TYPE.
SQ SEQUENCE 1131 AA; 129008 MW; 52C97388D1B6D04F CRC64;

Query Match 7.7%; Score 483; DB 1; Length 1131;
Best local Similarity 20.8%; Pred. No. 6,4e-29;
Matches 258; Conservative 204; Mismatches 560; Indels 220; Gaps 46;

```

```

QY 2 ENQKTVTPGPGLYGYACRVEDLDEISFLAARSTDSIALPLMRNLTVKPTSS 61
Db 3 DDETVVASDICTAAWLLYLPKEOKLEILETTLSLMKRRSVSPLLLNLTENDPEPT 62
QY 62 LAVVGARTTGAGAGITLKTTHSHFPVSFVHPGKGVLPSSA-ANLIPACAAAPRP 120
Db 63 V-----KPIINIGTIVITKITSFMPVCFPPFGTIVFLKFAADPHONIKLCKQTEKK 115
QY 121 GFSKCGPPVDGAVETTG-AEITRLEPEENTILYVLTALFEAVFMNVFLHYGLD 179
Db 116 NL--QEFVGNKNSVDICKICHSVGKNADVLCHIVGNGFELLF-----AGLL 164
QY 180 I-----VHINHDP-VIKIPLPVQVLFMDVNNLVDPNTHRSIGEGVYTPPYNT 231
Db 165 IPCVEEOIQVQVBCIAIKIPLYSATLFESEETICT-----DTCTEIDENGYPAP 215
QY 232 GLCHLHDCVIAPMVAVALRVNVAVARGAHLAFDENHRCVLPDITVYFQSSSGT 291
Db 216 QISEVLFYLLFTSMGMLLKHNNLTLELIKAGLKQPIQDTPQVLAHPKTY----- 265
QY 292 TTARGARNDVNSTSKRSPSGGFEERLASMADTLAAEYIN-----TGIEETP-T 344
Db 266 ---HGIPGOKLSPIEKD-----HMLVD-AVITETITSYAEVLDIYENQIM 310
QY 345 DIKEMPMFIMEGTLPLRLNAGSYTARVAGYIGAMVSPNSALYLTVEEDSGMTEAKDG 404
Db 311 NFEWPIIKSAETHEEKIVELKRLHLSHVAALVFAANSILYSNLAITSNK----- 365
QY 405 PGPSFN-----RPFQAPGHLAANPQTDRODH-VLSOSTGSSNTPESVDYIALIC 454
Db 366 --QAFNSAITQETLIRISQNSLSTNEDFYNDARKLIKNSSPKEDKSAFAPLAVAC 423
QY 445 GFGAPLIRLLFLEKRDCAAFITGHDALKYVGTGDSLEPGLCKKHTRPVANHITVH 514
Db 424 ATCPQILSHIIMNNRNSIYNTNCGNSEIYNHIVNC--SSNLCEFCGCKCHSIGTAL 481
QY 515 RLNRMPRQOATROPPIGVGTNNQSYSDCPDIGNAPVYLIIRKPGQUTAAKATMDITY 574
Db 482 RINSRLPQISKTTRKEETIVTMTESRFYADVGLSGFKKGNEKDKMEQUTPSIDRF 541
QY 575 KATLERFLIDBOBRLLDRAPCSSBGLSYIV-DHPYTR-----RIIDTLARAIKQ 626
Db 542 K-----FLGMIDY-----CKKNLIDAIIGEDMNPSONDFVNMINDLIQCLIEA 588
QY 627 TIGEMKVLVETRDYKIREGSEAIHSMALTEDPYSACFPTNLVYKRLHAAVQDALS 686
Db 589 VSKCISMRTQ--TSRQIENCLQSFNIDTTPLSLAFSPFYTYKYKILIVLQNLAL- 645
QY 687 QCHGVFYGOQVE-----GRNFRNOFQVLLRRRFDVLFNGCFITRSITVTLSFGPVS 738
Db 646 ---IIGTVVDDPCTGNLISKMLMOQYGLYGAFYNSHFKKGLNKKTVKI----- 694
QY 739 APPPTLQD-----DAVAGRTFQDGLARVSVEVIRDIHVKNRVVFSQNGCTLSAAR 789
Db 695 ASNVDMQYIDPFLNFSGAKTAKTSIQAKLCRLSMQCLDRVKKRPNNKKNQTN-- 751
QY 790 ARLVGLASAVQROEKRYDMLGALGFLKQFHGLFPRGPPNPKSPNPQWFTLLQBNQ 849
Db 752 -----PFFKVKYQKKNNPLSGCISFLFVYHERLF-----PNLKISCLR-FWQRIILNN 798
QY 850 MP--ADKLTHEEITT IAAVKRTEYEAALNFINLPTCIGELAOFMANLIKCYCHSV 906
Db 799 MPKTTIDIGNVELMKSFTEKTFKVTNSVDEIDLLDIOPECLLSFEYFHNKLSVLQYK 858
QY 907 YL--INTITSIIIGARRPRQSSVLMWIRKQV--SAAUJETAKALIEKTEENLPLMTWT 962
Db 859 YLVSIAHALSKIV---PQNPMLFPVFLKHPFTSSQGVYVMHAKIV--GNGIKLHPMTA 912
QY 963 AFTSTHLYRAANMORPVVVLGISISKYHGAAGNN--RVFOAGNKGSGINGKKNVCP- 1016
Db 913 SLTRKPNFGSIFTRGRIIIPGLMIKRFVSAVRDYFFHFGQIGWLAGSGVIRNLNPPSSGL 972

```

```

01 1017 FETDERKRFETLACVJGCEVIGCHSSGNGEETLSD--GVKGIH-----VSCGAMV 1066
02 974 GQDFEHR-GRKEVIAIK-----LEGIH-VKKYKREAIYVEVEYIRCKVNLIEELSNVNE 1026
03 1067 LAIVATVYAVAVARAGQAHAFDMLDLSLTD-ELFANLE-----ELHDIOTLETP 1116
04 1027 ELIIAEVNM-KDRDSKPLMDMLFYVDORFPIAKSNMKGHTLMDIANVHDFSLSTLISV 1083
05 1117 W-----TVEALEVAKIIDEKTTAGDDEPT 1142
06 1004 EEEVVEDSNAIYDSEELVEVNEGEFGLCKEETEHENEEPS 1125
07
08 RESULT 15
09 ENH1 SCVVC STANDARD; PRT: 1160 AA.
10 AC P14215;
11 01 JAN 1990 (rel. 13, created)
12 01 APR 1994 (rel. 25, last sequence update)
13 16 OCT 2001 (rel. 40, last annotation update)
14 Major DNA binding protein (MGBP).
15 1167 OR DBP.
16 Simian cytomegalovirus (strain Colburn).
17 Genus: GCMV.
18 Genus: GCMV.
19 Genus: GCMV.
20 Genus: GCMV.
21 Genus: GCMV.
22 Genus: GCMV.
23 Genus: GCMV.
24 Genus: GCMV.
25 Genus: GCMV.
26 Genus: GCMV.
27 Genus: GCMV.
28 Genus: GCMV.
29 Genus: GCMV.
30 Genus: GCMV.
31 Genus: GCMV.
32 Genus: GCMV.
33 Genus: GCMV.
34 Genus: GCMV.
35 Genus: GCMV.
36 Genus: GCMV.
37 Genus: GCMV.
38 Genus: GCMV.
39 Genus: GCMV.
40 Genus: GCMV.
41 Genus: GCMV.
42 Genus: GCMV.
43 Genus: GCMV.
44 Genus: GCMV.
45 Genus: GCMV.
46 Genus: GCMV.
47 Genus: GCMV.
48 Genus: GCMV.
49 Genus: GCMV.
50 Genus: GCMV.
51 Genus: GCMV.
52 Genus: GCMV.
53 Genus: GCMV.
54 Genus: GCMV.
55 Genus: GCMV.
56 Genus: GCMV.
57 Genus: GCMV.
58 Genus: GCMV.
59 Genus: GCMV.
60 Genus: GCMV.
61 Genus: GCMV.
62 Genus: GCMV.
63 Genus: GCMV.
64 Genus: GCMV.
65 Genus: GCMV.
66 Genus: GCMV.
67 Genus: GCMV.
68 Genus: GCMV.
69 Genus: GCMV.
70 Genus: GCMV.
71 Genus: GCMV.
72 Genus: GCMV.
73 Genus: GCMV.
74 Genus: GCMV.
75 Genus: GCMV.
76 Genus: GCMV.
77 Genus: GCMV.
78 Genus: GCMV.
79 Genus: GCMV.
80 Genus: GCMV.
81 Genus: GCMV.
82 Genus: GCMV.
83 Genus: GCMV.
84 Genus: GCMV.
85 Genus: GCMV.
86 Genus: GCMV.
87 Genus: GCMV.
88 Genus: GCMV.
89 Genus: GCMV.
90 Genus: GCMV.
91 Genus: GCMV.
92 Genus: GCMV.
93 Genus: GCMV.
94 Genus: GCMV.
95 Genus: GCMV.
96 Genus: GCMV.
97 Genus: GCMV.
98 Genus: GCMV.
99 Genus: GCMV.
100 Genus: GCMV.

```

[illegible]

```

QY 1049 TISDQVKG11VSCGAMVOL-----A1YATVRAVAGARAOHMAFDMLSLTDD- 1095
Db 1020 VLFDE---DIKRVMAALDSENLDVDPPELMAY-----ETLSTREEIPERDDVLEFVGGC 1072
QY 1096 EFLARDLEELHDOIIQTLTPWTVESALEAVKILDEKTTAGDEPTPTNLA FNDSCEPSH 1155
Db 1073 QAVVADSLMEKFSRLQPMGVDDPFSI---VNLOQVLDSPRCCGGGGEV-----H 1117
QY 1156 DTSNVLNIGSNISGSTVPGLK-RPPED--DELFDSGIPIKHGN 1199
Db 1118 DLSALFTAASGEAVGNSV--GLNARGGEHAFFE--DGGILPAKRGRL 1160

```

Search completed: March 28, 2003, 13:36:03  
 Job time : 37 secs

